The Journal of INDO-EUROPEAN STUDIES

Volume 43, Numbers 3 & 4 **Fall/Winter 2015**

Proto-Indo-European-Uralic Comparison from the Probabilistic Point of View

Alexei Kassian,[1] *Mikhail Zhivlov,*[2] and *George Starostin*[3]

In this paper we discuss the results of an automated comparison between two 50-item groups of the most generally stable elements on the so-called Swadesh wordlist as reconstructed for Proto-Indo-European and Proto-Uralic. Two forms are counted as potentially related if their first two consonantal units, transcribed in simplified consonantal class notation (a rough variant of the Levenshtein distance method), match up with each other. Next to all previous attempts at such a task (Ringe 1998; Oswalt 1998; Kessler & Lehtonen 2006; Kessler 2007), our automated algorithm comes much closer to emulating the traditional procedure of cognate search as employed in historical linguistics. "Swadesh slots" for protolanguages are filled in strict accordance with such principles of reconstruction as topology (taking into consideration the structure of the genealogical tree), morphological transparency, typology of semantic shifts, and areal distribution of particular items. Altogether we have counted 7 pairs where Proto-Indo-European and Proto-Uralic share the same biconsonantal skeleton (the exact same pairs are regarded as cognates in traditional hypotheses of Indo-Uralic relationship). To verify the probability of arriving at such a result by chance we have applied the permutation test, which yielded a positive result: the probability of 7 matched pairs is equal to 1.9% or 0.5%, depending on the constituency of the consonantal classes, which is lower than the standard 5% threshold of statistical significance or even lower than the strong 1%

[1]Institute of Linguistics of the Russian Academy of Sciences / School for Advanced Studies in the Humanities of the Russian Presidential Academy of National Economy and Public Administration; a.kassian@gmail.com
[2]Department of Comparative Linguistics of the Russian State University for the Humanities / School for Advanced Studies in the Humanities of the Russian Presidential Academy of National Economy and Public Administration, zhivlov@gmail.com
[3]Department of the History and Philology of the Far East of the Russian State University for the Humanities / School for Advanced Studies in the Humanities of the Russian Presidential Academy of National Economy and Public Administration; gstarst1@gmail.com

level. Standard methodology suggests that we reject the null hypothesis (accidental resemblance) and offer a more plausible explanation for the observed similarities. Since the known typology of language contacts does not speak in favor of explaining the observed Indo-Uralic matches as old lexical borrowings, the optimal explanation is seen in the hypothesis of an Indo-Uralic genetic relationship, with the 7 matching pairs in question representing archaic retentions, left over from the original Indo-Uralic protolanguage.

1. Previous studies

The hypothesis of a genetic relationship between the Indo-European and Uralic languages was suggested already in the 19th century (Thomsen 1869); later supported by several distinguished comparative linguists (e.g., Collinder 1934), it arguably reached its highest degree of etymological elaboration in V. M. Illich-Svitych's work on "Nostratic" linguistics (Illich-Svitych 1967; 1971–1984), and was further developed by certain scholars of the Leiden school (e.g., Kortlandt 2010; Kloekhorst 2008a). Regardless of any differences in scope and quality, all the attempts to produce convincing etymological evidence for "Indo-Uralic" have so far been met with widespread skepticism. However, since at least a certain amount of similarities between the two language families is undeniable, "Indo-Uralic" has also been a relatively popular subject of investigation under various probabilistic approaches, usually as a convenient case study of estimating the probabilities of determining genetic relationship on a deep chronological scale. The most important publications on this issue are as follows:

1) Ringe 1998. The author compares two 100-item wordlists of Proto-IE and Proto-Finno-Ugric (not Proto-Uralic) based on a very specific algorithm that considers recurrent sound correspondences between two lists rather than simple phonetic similarity (besides Ringe 1998, see Ringe 1992 and especially Baxter & Manaster Ramer 1996 for details). Ringe's conclusion is that comparison of initial root consonants yields a **positive** result, with $p = 0.05$, although comparison of second consonants is unsuccessful, with $p = 0.18$ or 0.5 (Ringe 1998: 174, 177). There are two principal issues with Ringe's stud-

ies. First, the two proto-lists were specially designed to maximize resemblances with each other (not all possible synonyms were included in the corresponding slots, and, on the other hand, the word 'ice' was intentionally added to the list, Ringe 1998: 173). Second, this positive probabilistic result is mostly based on coincidences that are unlikely to reflect true etymological sound correspondences, such as IE **d* — Finno-Ugric **p*, etc. (Ringe 1998: 167).

2) Oswalt 1998. The author compares 100-item wordlists of Modern Russian (a representative of the IE family), Finnish, Hungarian and Nenets (representing the main groups of the Uralic family). Shift test was performed (a kind of permutation test with X–1 trials for X-item word lists, i.e., 99 trials in our case); two compared forms are considered equal if any two consonants in one form coincide with any two consonants in the other form (with different degrees of similarity), regardless of their position in the word (for instance, an initial consonant in one form can match a medial consonant in the compared form). All three IE-Uralic language pairs offer **positive** results; normally more strong criteria of phonetic similarity decrease the probability p of chance coincidences. The best result for Russian-Finnish and Russian-Hungarian comparisons is $p < 0.001$. The best result for Russian-Nenets comparisons is $0.01 < p < 0.05$. Earlier IE-Uralic studies by the same author (Oswalt 1970; Oswalt 1991) had likewise yielded a positive signal.

3) Kessler & Lehtonen 2006. The authors offer a multilateral comparison of 100-item wordlists for 11 IE and 4 Uralic languages. Permutation test was performed, for which crude Levenshtein distances (based on place of articulation only) between initial root consonants were measured. A **negative** result was obtained: $p = 0.45$, i.e., the probability that the matches between IE and Uralic are due to chance is too high (45%). Kessler & Lehtonen also report that additional bilateral comparisons between individual IE and Uralic languages were performed as well,

but none of the IE-Uralic pairs yielded a positive result either, i.e., $p > 0.05$ in all the cases.

4) Kessler 2007. The same multilateral procedure is applied to the same IE and Uralic data as in Kessler & Lehtonen 2006, but with seven different methods of measuring distances between compared forms. Out of these seven tests, five yielded negative results ($p > 0.05$), but two tests produced **positive** results, suggesting that lexical matches between IE and Uralic are not chance coincidences. Both of these successful tests are based on Dolgopolsky consonant classes (for which see §2.2). Multilateral IE-Uralic comparison for initial root consonants yields $p = 0.04$; multilateral IE-Uralic comparison based on primitive alignment of consonantal skeletons yields $p = 0.02$.

Despite the fact that the majority of the aforementioned tests yielded positive results, these studies face some serious difficulties. The main theoretical problem is that the proposed tests do not imitate the real work of historical linguists and do not reproduce the traditional comparative procedure[4] (for additional conceptual flaws of the approaches in Oswalt 1998 and especially Ringe 1998, see Baxter 1998). The practical problem is that the input lexical wordlists do not meet the strong lexicographic criteria accepted, e.g., in the *Global Lexicostatistical Database* project (for which see, e.g., G. Starostin 2013, Kassian 2015).

2. Method

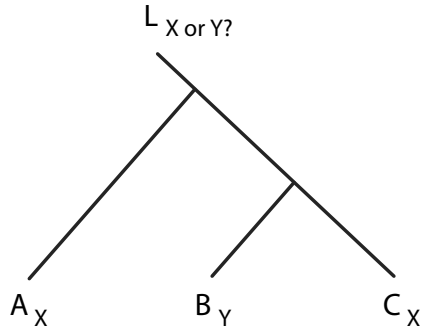
2.1. Principles of semantic reconstruction. The following several principles should be observed in proper etymological reconstruction of protolanguage morphemes as entities that are endowed with a specific meaning (e.g., 'hand', 'red', 'to come') rather than merely assigned to a general

[4]For instance, Ringe's (1998) algorithm approves such odd sound correspondences as IE **d* — Finno-Ugric **p* or IE **h*$_2$ (i.e., zero) — Finno-Ugric **l*. On the other hand, Oswalt (1998) treats two forms as cognates if they share at least two matching consonants in *any* position, e.g., Russian *korʲinʲ* ⟨корень⟩ 'root' appears to be cognate to Hungarian *ɖökeːr* ⟨gyökér⟩ 'root' (*kr* = *kr*), whereas Russian *nogəcʲ* ⟨ноготь⟩ 'fingernail' is cognate with Finnish *künsi* ⟨kynsi⟩ 'fingernail' (*ng* = *kn*), etc.

semantic field (e.g., 'part of limb', 'kind of bright color', 'to move').

1) *Topological principle.* The configuration of the genealogical tree of a linguistic family is important for semantic reconstruction. In the situation when reflexes of a proto-word have different meanings in daughter languages, one of the strongest criteria for reconstructing the original meaning is the topological one. Let us envision the following genealogical tree where *L* is a proto-language and *A*, *B*, *C* are its daughter languages.

Figure 1. A language family tree.



A certain word in *A* means 'X', its etymological cognate in *B* has a different meaning 'Y', whereas its cognate in *C* also means 'X'. In the absence of additional evidence, the likeliest solution is that, in *L*, this word denoted 'X' rather than 'Y', since, according to general scientific principles, we should prefer the most economic scenario (one semantic shift 'X' > 'Y' in *B* vs. two independent shifts 'Y' > 'X' in *A* and *C*).

2) *External etymology principle.* If there are two or more equally probable lexical candidates for the status of the proto-term in question, the one root for which a better external (i.e., outside the examined group of languages) etymology is found, especially if the external cognate preserves the same meaning, has the advantage.

3) *Internal etymology principle.* If there are two lexical candidates for the status of the proto-term in question, and one of them is morphologically primary, whereas the other is transparently derivative (for example, 'moon' ← 'to shine',

rivative (for example, 'moon' ← 'to shine', 'green' ← 'grass' and so on), the first term has an advantage. The same applies to the situation when one of the competing terms possesses some etymological cognates (either internal, that is, within the same language group, or external), whereas the second term is etymologically isolated. In this case, the first term has an advantage, since the second one is more likely to represent a potential loanword.

4) *Semantic plausibility principle.* If there are two lexical candidates for the status of the proto-term in question, and both of them have assorted cognates in related lects, the assumed semantic shifts must be examined and weighted. Although many common semantic shifts are bilateral, in some cases only one specific direction of semantic development is typologically probable. For instance, the shift 'green' ↔ 'grass' is bilateral (there are plenty of typological confirmations for both directions), but in the pair 'moon' and 'to shine' only the development 'to shine' → 'moon' has been attested, never vice versa.

5) *Areal effect exclusion principle.* If the term in question has a specific areal distribution, i. e., is attested in several neighboring lects, and we have strong evidence that these lects are in contact and influence each other, such a word can represent a late introduction, which has spread as an interdialectal loan.

See G. Starostin 2013: 153–183; Kassian 2013b, where these and additional, less important, principles are explained in detail and illustrated with numerous examples from a diverse selection of language families.

In addition to observing these principles as rigorously as possible (all except for the external etymology principle, which is completely excluded since the genetic relationship between IE and Uralic has not yet been proven), it is also our intention to maximally formalize the reconstruction procedure for Proto-IE and Proto-Uralic lexical items. This means that it is allowed and common for more than one proto-form to be assigned to the same "proto-slot"

where the evidence remains ambiguous even after the application of all the abovementioned principles. The relative "champion" in this respect is the Proto-Uralic slot #78 'smoke', for which we are required to postulate six technically synonymous Proto-Uralic stems (this does not imply the presence of six different terms for 'smoke' in Proto-Uralic, but simply our inability to pinpoint the most probable one). Naturally, such solutions increase the probability of the observed IE-Uralic matches being due to chance. Nevertheless, we believe that in a test like this, it is important to adhere to the most rigorous, subject-independent methods as the first step of probabilistic comparison.

2.2. Principles of cognation marking. For the purposes of automatic match-making between IE and Uralic protoforms that could be suspected of being related, we use a formal algorithm, based on phonetic similarity. The two most popular approaches to the automatic establishing of quasi-cognate word pairs between fixed wordlists are (a) Levenshtein distances and (b) consonant classes. Actually, the method of consonant classes may be considered a crude variation on the measurement of Levenshtein distances.[5] Since we are not aware of any publications that would demonstrate that consonant classes yield significantly less reliable results than Levenshtein distances, and the method itself is relatively simpler in terms of practical application, below we will rely exclusively on consonant classes for evaluation of lexical data.

The method of consonant classes was proposed by A. Dolgopolsky in 1964 (English version: 1986) and successfully tested by various authors on the data of various languages of Eurasia and Africa.[6] This method implies that the phonetic alphabet used in our studies can be divided into several non-intersecting subsets (classes) in such a way that phonetic mutations between the sounds of one class during natural language development are typologically more

[5]The algorithm can be described as measuring Levenshtein distance (LD) between two forms, but subsequently identifying only pairs with LD = 0 as hypothetical etymological cognates, whereas pairs with LD > 0 are treated as too expensive.

[6]E.g., Baxter 1995; Baxter & Manaster Ramer 2000; Kessler 2007; G. Starostin 2008, 2013; Turchin et al. 2010.

normal than mutations between sounds that belong to different classes (with individual exceptions that may be overlooked in a rough application of the method).[7] The several distinct classes that have been currently agreed upon in the *Global Lexicostatistical Database* project are as follows:[8]

Table 1. Basic GLD consonant classes

P-class (labials): p b ɓ β f v…
T-class (dentals): t d ɗ θ ð…
S-class (sibilant affricates & fricatives): c ʒ č ǯ s z š ž…
Y-class (palatal glides): y…
W-class (labial glides): w ʍ…
M-class (labial nasals): m ɱ…
N-class (non-labial nasals): n ɳ ɲ ŋ…
Q-class (lateral affricates): ƛ …
R-class (liquida): r ɾ l ɭ ɬ ɮ…
K-class (velars & uvulars): k g x ɣ q χ ʁ…
zero-class or H-class: ħ ʕ ʜ ʢ ʔ h ɦ ʔ and any vowels.

This simplified transcription system (*P T S Y W M N Q R K H*) is sufficient to encode all wordforms or morphemes of any natural language that is included in comparison. Elements of the *H*/zero-class and such features as coarticulation, prosody, phonation are notably deleted from the structure, with the exception of word-initial and word-final vowels and laryngeals which are coded as *H*. Thus, such hypothetical forms as *tasa* and *dʰüʒo* are both encoded as *TSH*; *alaq* and *ʔärx* = *HRK*; *na* and *ŋoʔ* = *NH*; *pkʰot* and *baqʼaθ* = *PKT*; *wahat* and *ʍad* = *WT*. Another exception requires non-initial *Y* and *W* (weak glides) to be treated as *H*, i.e., they are deleted in the medial position and coded as *H* in the final position, thus *ka*, *kay*, *kawa* = *KH*, whereas *kat* and *kayat* = *KT*. As follows from the above, two forms

[7]Although typology of sound change still remains an insufficiently well explored area, cf. Brown et al. 2013; Kümmel 2007 for some progress.

[8]http://starling.rinet.ru/new100/sound.pdf [accessed 20.04.2014]. All linguistic data in the present article are encoded in the unified transcription system of the *Global Lexicostatistical Database* project, which is generally based on the IPA alphabet, with just a few specific discrepancies (http://starling.rinet.ru/new100/UTS.htm). The only transcriptional exception introduced in the present paper due to the editor's request is the ring below symbol, which we accept for syllabic sonants in Proto-Indo-European forms (l̥, m̥, n̥, r̥) instead of the IPA vertical stroke below (l, m, n, r).

from compared languages possessing identical simplified transcriptions generally have a better a priori chance to represent true etymological cognates than forms whose simplified transcriptions differ.

We have subsequently encoded wordlists for Proto-IE and Proto-Uralic in this way, whereupon the StarLing software (S. Starostin 1993/2007; Burlak & Starostin 2005: 270 ff.) has permitted automatic marking of all the lexical matches. Two forms are marked as cognates if they share the same first two consonants in their simplified transcriptions, e. g., the words for 'water': IE **wed-* (*WT*) = Uralic **weti* (*WT*). On the other hand, IE **kʸerd-* 'heart' (*KR*) ≠ Uralic **sʸäðʸä* id. (*ST*).

Additionally, we have also performed the same procedure with a more detailed set of consonant classes, in which (a) the S-class is divided into a S-class proper (sibilant fricatives: s z š ž…) and a ʒ-class (sibilant affricates: c ʒ č ǯ…); (b) the R-class is divided into a R-class proper (r ɾ…) and a L-class (l ɭ ɬ…)[9]. The full modified list of classes now has the following look (new classes are marked with an asterisk *): Table 2.

Table 2. More precise consonant classes

P-class (labials): p b ɓ β f v…
T-class (dentals): t d ɗ θ ð…
S-class (sibilant fricatives): s z š z…
*Ʒ-class (sibilant affricates): c ʒ č ǯ…
Y-class (palatal glides): y…
W-class (labial glides): w ʍ…
M-class (labial nasals): m ɱ…
N-class (non-labial nasals): n ɳ ɲ ŋ…
Q-class (lateral affricates): ƛ …
R-class: r ɾ…
*L-class: l ɭ ɬ…
K-class (velars & uvulars): k g x ɣ q χ ʁ…
zero-class or H-class: ħ ʕ ʜ ʢ ʔ h ɦ ʔ and any vowels.

2.3. Permutation test. We perform the so-called permutation test as outlined and tested by W. Baxter & A. Manaster

[9]At least in the case of *R* and *L*, this is logical since shifts between *l*-like and *r*-like sounds are atypical for IE and Uralic languages (except for the Indo-Iranian group) as well as for neighboring linguistic areas.

Ramer (2000) and some other authors.[10] Below, Proto-IE-Uralic lexicostatistical matches are evaluated with the help of Baxter & Manaster Ramer's (2000) algorithm, which has been recently incorporated as an optional plug-in into the StarLing software.

The principle of the permutation test is simple and elegant. For two bi-unique and uniformly transcribed wordlists with *X* lexical phonetic matches, the test begins by randomly reshuffling one of the lists, checking the number of matches for each new configuration. If the number of random configurations is sufficiently large, it becomes possible to establish a statistical norm for the average amount of expected accidental matches and, second, to calculate the probability of *X* and more than *X* matches between our original lists.

As noted above, in our study, Proto-IE and Proto-Uralic 50-item wordlists have been transcribed according to the above-described simplified notation of consonant classes. Two forms are counted as constituting a positive pair if the first two consonants (*CC*) of the Proto-IE form are identical to those of the Proto-Uralic form. When the same slot is occupied by several synonyms (a normal situation in our case), we compare all possible pairs between two languages: if there is at least one matching pair, the whole slot is treated as positive. E.g., the slot #39 'to hear' is filled with **k^{y}lew-* for IE, but with two equiprobable candidates **kuwli-* and **yünti-* for Uralic; the pair **k^{y}lew-* — **kuwli-* is positive (*KL* = *KL*), whereas **k^{y}lew-* — **yünti-* is negative (*KL* ≠ *YN*), so the whole slot #39 'to hear' is positive.

1,000,000 random (strictly speaking, pseudorandom) trials have been performed in each case described below.

3. Data

The relationship between languages is generally demonstrated with common etymologies of root mor-

[10] The general idea goes back to Oswalt 1970; for a further overview see McMahon & McMahon 2005: 66–68. See also Justeson & Stephens 1980; Baxter 1995; Kessler & Lehtonen 2006; Kessler 2007; Dunn & Terrill 2012 for application of the permutation test to lexical lists of specific languages. A very similar bootstrap procedure was described and successfully applied to various languages of Eurasia by Turchin et al. (2010).

phemes that are attested in the basic vocabulary (Campbell & Poser 2008: 4; Burlak & Starostin 2005: 7–24). For practical purposes, it is relatively reasonable and convenient to equate the "core" of the basic vocabulary with the so-called Swadesh 100-word-list. The Swadesh list is not homogenous, and it is commonly acknowledged that its entries possess different average degrees of stability (see S. Starostin 2007a for such an arrangement, based on a typological survey of language families of the Old World, further see G. Starostin 2010). The older the relationship between languages, the fewer "unstable" lexical matches on the Swadesh list one would expect to witness; this means that, for the purposes of verifying hypotheses of such old relationship, it makes good sense to reduce the overall list to a smaller number of "generally stable" items.

For the purposes of IE-Uralic comparison, we employ a lexical subset of the Swadesh wordlist that is reduced to 50 basic concepts and currently serves as one of the universally applicable standards in the Moscow school of historical linguistics. The list comprises the top half of S. Starostin's stability rating except for 9 items — 'this', 'that', 'liver', 'fish', 'neck', 'breast', 'full', 'to stand', 'to give' — which were excluded for various reasons and replaced with 'to kill', 'foot', 'horn', 'to hear', 'meat', 'egg', 'black', 'head', 'night', see G. Starostin 2010 for details. To be precise, the list that is used in the present paper differs in one slot from the 50-item wordlist of G. Starostin 2010: the IE term for 'louse' is all but impossible to reconstruct, since the Narrow IE expressive root **luː-(s)-* (Baltic, Germanic, Celtic; further perhaps Old Indic *yuːka-* and Slavic **vŭšĭ*) regularly corresponds to Tocharian B *luwo* 'animal', implying that the original meaning of **luː-* could be 'animal',[11] whereas Anatolian and Tocharian words for 'louse' so far remain undocumented; as for Ancient Greek ⟨φθείρ⟩ 'louse', it looks like a substrate Pre-Greek word.[12] Because

[11] For the euphemistic semantic shift 'animal' > 'louse', cf., e.g., Middle Irish ⟨míl⟩ and Modern Irish ⟨míol⟩ 'animal', which can also be used specifically in the meaning 'louse'.

[12] Apparently it was secondarily contaminated with Greek ⟨φθείρω⟩ 'to ruin, destroy', Beekes 2010: 1569 (etymological derivation 'to destroy' > 'louse' seems very odd from the typological point of view).

of this, we have replaced 'louse' with 'liver', which actually possesses a very high rank of stability (25) and was expelled from the 50-item wordlist due to extra-linguistic reasons: words for 'liver' frequently remain non-documented in various glossaries and wordlists of the world's languages.

Thus, the complete 50-item wordlist used in the present paper is as follows (word numbering is not continuous, since this is an excerpt from the full 100-itemlist; index figures designate the relative index of stability: 'we$_1$' is the most stable word, 'night$_{50}$' is the least stable one): Table 3.

Table 3. Semantic slots of the 50-item wordlist

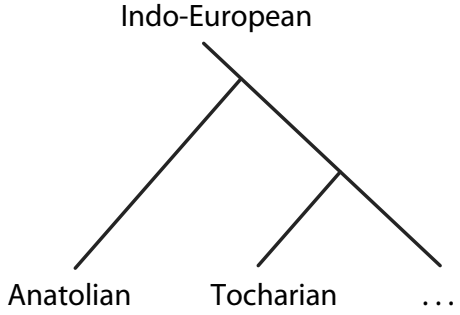
2.	ashes$_{38}$	37.	hand$_{11}$	63.	one$_{21}$
6.	bird$_{33}$	38.	head$_{9}$	65.	rain$_{39}$
8.	black$_{48}$	39.	hear$_{45}$	78.	smoke$_{36}$
9.	blood$_{20}$	40.	heart$_{14}$	80.	star$_{40}$
10.	bone$_{34}$	41.	horn$_{44}$	81.	stone$_{9}$
13.	fingernail$_{19}$	42.	I$_{3}$	82.	sun$_{35}$
17.	die$_{13}$	43.	kill$_{42}$	84.	tail$_{26}$
18.	dog$_{16}$	46.	leaf$_{41}$	87.	thou$_{5}$
19.	drink$_{15}$	48.	liver$_{25a}$	88.	tongue$_{8}$
20.	dry$_{24}$	53.	meat$_{46}$	89.	tooth$_{22}$
21.	ear$_{32}$	54.	moon$_{18}$	90.	tree$_{37}$
23.	eat$_{25}$	56.	mouth$_{31}$	91.	two$_{2}$
24.	egg$_{47}$	57.	name$_{10}$	94.	water$_{28}$
25.	eye$_{4}$	59.	new$_{23}$	95.	we$_{1}$
28.	fire$_{7}$	60.	night$_{50}$	96.	what$_{12}$
31.	foot$_{43}$	61.	nose$_{29}$	98.	who$_{6}$
36.	hair$_{27}$	62.	not$_{30}$		

For semantic specification of the Swadesh items and the general principles of the compilation process see Kassian et al. 2010.

In the above discussion on the principles of semantic reconstruction it has already been mentioned that the correct topology of the genealogical tree is a key condition for the reconstruction of ancestral states (this is a crucial difference between historical linguistics and modern molecular biology). Although some particular details of both the IE and the Uralic trees are still debatable, their main nodes are generally uncontroversial.

We proceed from the following IE tree: Fig. 2.

Figure 2. Principal phylogeny of the IE family



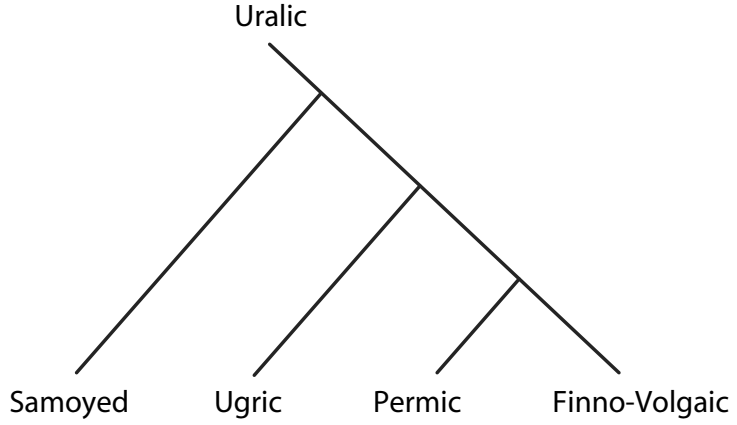
The early split of Anatolian is accepted by the majority of Indo-Europeanists, see Blažek 2007 for overview. Tocharian as the next splitting branch is also widely accepted, see, e.g., Gamkrelidze & Ivanov 1984/1995: 350, Winter 1996, Anthony & Ringe 2015, the same topology with the Anatolian and Tocharian outliers is suggested by the majority of published formal phylogenies: e.g., Ringe et al. 2002, Nakhleh et al. 2005, Blažek 2007, and some others.

Such crude topology is sufficient for the majority of Swadesh items that are relevant for our paper. In any case, it must be stressed that our general approach is quite "liberal", i.e., if there is any suspicion that the given proto-root cannot be explained as a local innovation, we prefer to include it as a synonym in the Proto-IE list.[13]

We proceed from the following Uralic tree: Figure 3.

[13] E.g., for the IE slot 'fire', we accept two synonyms: **pex-wer-* (Anatolian, Tocharian, Greek, Armenian, Germanic, Italic) and **n̥g-n-i-* (Indic, Baltic, Slavic, Italic). The latter stem is certainly weaker in terms of overall distribution; nevertheless, it cannot be formally excluded, since there is no evidence of Indic, Baltic, Slavic and Italic constituting a phylogenetic cluster that does not include Germanic.

Figure 3. Principal phylogeny of the Uralic family



Such a topology is widely accepted by Uralists, see, e.g., Collinder 1960: 11, Sinor 1988, Sammallahti 1988, Napolskikh 1997: 256 ff. (although cf. criticism in Salminen 2002).[14]

We use a unified phonetic transcription for all IE and Uralic forms involved in our tests.[15] The following transcriptional peculiarities should be noted:

1) The so-called laryngeal phonemes in IE proto-forms are irrelevant for our purposes: e.g., the sequences *op-* and *hop-* are both transcribed as *HP*, *əl-* and *hl-* are both transcribed as *HL*, *paːs-* and *pahs-* are both transcribed as *PS*. Since even the authors of the present paper hold somewhat different opinions on the subject, we add laryngealistic proto-forms in parentheses.
2) When relevant Anatolian data are available, we introduce the velar fricative **x* (> cuneiform ⟨ḫ, ḫḫ⟩) into the IE reconstruction. E.g., IE 'bone' is reconstructed as **xost-* (i.e., *KST* in the simplified transcription), not as **h*$_2$*est-* (*HST*).[16]

[14] In a recent paper Syrjänen et al. 2013, a new lexicostatistical tree for Uralic has been proposed, according to which Permic splits off prior to Ugric. However, since the authors have not disclosed the actual lexical dataset used to produce the tree, their results remain unverifiable.

[15] As mentioned above, the transcription system employed in the *Global Lexicostatistical Database* project is mostly based on the IPA alphabet: http://starling.rinet.ru/new100/UTS.htm [accessed 20.04.2014].

[16] Despite the Indo-Europeanist term "laryngeals", the Hittite and Luwian phonemes covered under the cuneiform signs for ⟨ḫ, ḫḫ⟩ most likely represented the velar-uvular fricative *x* ~ *χ*. The main arguments for this

3) An additional IE fricative *θ should probably be reconstructed for the correspondence Hittite *s* / Luwian *t* / Narrow IE *0*, as proposed in Ivanov 2001: 133; 2009: 5 and (independently) in Kassian & Yakubovich 2013: 22. In our list, this affects two items: #13 'fingernail' & #25 'eye'. In both cases, we double the reconstructed forms with and without such a fricative, e.g., 'eye' is reconstructed as two synonyms **θok^{w}-* ~ **ok^{w}-* (i.e., *TK* ~ *HK* in the simplified transcription).
4) Phonetic shapes of the reconstructed Uralic forms generally follow Sammallahti 1988 with certain emendations from Zhivlov 2014. Uralic Phonetic Alphabet distinguishes between full mid central unrounded vowel ⟨ɐ̟⟩ and "reduced" vowel ⟨ə⟩. In our transcription UPA ⟨ɐ̟⟩ is replaced by *ə*, while UPA ⟨ə⟩ is replaced by *ə̆*
5) Double asterisk ** is used for those Proto-IE & Proto-Uralic items which are technically reconstructed on the basis of isolated forms, only attested in one group of the corresponding family.

As a result, the following Proto-IE and Proto-Uralic 50-item wordlists have been obtained: Table 4 (table in MS Excel-format is available on-line as Supplementary materials; linguistic comments on each reconstructed form are available in the Appendix section below). Within one semantic slot, our computer algorithm treats two forms as synonyms if they are either assigned to two different cells or quoted within one cell, but separated with the tilde sign (~). Thus, e.g., for the Uralic slot 'tooth', three synonyms are analyzed: **piŋi* [*PN*] / ***timä* [*TM*] / ***simä* [*SM*] (for synonyms in the permutation test, see §2.3 above). Forms

---

statement are as follows: (a) the Hittite cuneiform signs for ⟨ḫ, ḫḫ⟩ were borrowed from Akkadian, where they denoted a sound that originated from the Proto-Semitic velar-uvular fricative **x* ~ **χ*; (b) attested Egyptian and Ugaritic transcriptions of Hittite forms rendered cuneiform ⟨ḫ, ḫḫ⟩ as *x* ~ *χ* rather than as laryngeal sounds (Patri 2009); (c) even though the phonetic shift *laryngeal* > *velar* is quite uncommon cross-linguistically, Lycian still had a *k*-like sound as an etymological counterpart of Hittite-Luwian ⟨ḫ, ḫḫ⟩, also suggesting a velar articulation of the original phoneme.

in parentheses (IE laryngealistic reconstructions) are ignored by the formal algorithm.

Table 4. Proto-IE and Proto-Uralic 50-item wordlists

#	Word	Indo-European	Uralic
2	**ashes**$_{38}$	*xaːs- (*h_2eHs-)	*kaðʸma
6	**bird**$_{33}$	*awi- ~ *xawi- ~ *wi- (*$h_2$ewi- ~ *$h_2$wei-)	**sʸarma **nVrmV- **mVntV
8	**black**$_{48}$	*kr̥s-no- *melə-n- (*melh$_2$-n-) *dʰengʷ-	**musi- **sʸäpːä *rapša **sʸɨnti *pekːVtːV
9	**blood**$_{20}$	*esx-r̥ (*$h_1$esh$_2$-r̥)	*weri **kɨmi
10	**bone**$_{34}$	*xost- (*$h_2$est-)	*luwi
13	**claw(nail)**$_{19}$	*θmogʰ- ~ *θŋ̥gʰ- ~ *mogʰ- ~ *nogʰ- ~ *ŋ̥gʰ- (*$h_3$nogʰ- ~ *h_3ŋ̥gʰ-)	*künči **kuta
17	**die**$_{13}$	*wel-	*kali-
18	**dog**$_{16}$	*kʸwon-	*penä *ämpV **wɨni
19	**drink**$_{15}$	*eːgʰʷ- (*$h_1$egʰʷ-)	*yɨɣi- ~ *ɨɣi-
20	**dry**$_{24}$	*saws-	*kosʸka
21	**ear**$_{32}$	*ows- ~ *xows- (*$h_2$ews- ~ *$h_2$ows-) **stom-	*pelyä **kawi
23	**eat**$_{25}$	*ed- (*$h_1$ed-)	*sewi- *imi-
24	**egg**$_{47}$	*oːwyo- ~ *xoːwyo- (*h_2oːwyo-)	*muna
25	**eye**$_{4}$	*θokʷ- ~ *okʷ- (*$h_3$ekʷ-)	*sʸilmä
28	**fire**$_{7}$	*pex-wr̥ (*peh$_2$-wr̥) *ŋ̥g-n-i-	*tuli

31	**foot**$_{43}$	*pod-	*yalka **ayi ~ **ali
36	**hair**$_{27}$	**ted- **top- *dek^{y}- *wel-	*ɨpti
37	**hand**$_{11}$	*g^{hy}es-r̥	*käti **VtV
38	**head**$_{9}$	*k^{y}er- *g^{h}eb^{h}-Vl- **xors-	*oywa
39	**hear**$_{45}$	*k^{y}lew-	*kuwli- **yünti-
40	**heart**$_{14}$	*k^{y}erd-	*s^{y}äð^{y}ä
41	**horn**$_{44}$	*k^{y}r-ew-r̥	**amti
42	**I**$_{3}$	*eg^{y}- (*h$_1$eg^{y}H-)	*mi-n
		*me- ~ *əme- (*h$_1$me-)	
43	**kill**$_{42}$	*g^{hw}en-	*wiði-
46	**leaf**$_{41}$	**porst- *b^{h}el-	*lVpV
48	**liver**$_{25a}$	*yek^{w}-r̥ **les- **pVn-	*mɨksa
53	**meat**$_{46}$	*mems-	*siwVlV **ɲVkVð^{y}V **aya ~ **að^{y}a
54	**moon**$_{18}$	*meːn- (*meh$_1$n̥s-) **or-	*kuŋi
56	**mouth**$_{31}$	*oy-es (*HoH-es-)	*s^{y}uwi **aŋi
57	**name**$_{10}$	*nom-n̥ ~ *lom-n̥ ~ *ənom-n̥ ~ *ələm-n̥ (*h$_1$nom-n̥ ~ *h$_1$n̥m-en-)	*nimi
59	**new**$_{23}$	*new-	*wuð^{y}i
60	**night**$_{50}$	*ksep-	*eyi **piyV ~ **piɣV

61	**nose**$_{29}$	*nas- **tid- **mVl-	*näri **pɨya ~ **pɨðʸa **ɲulV ~ **ɲuðV
62	**not**$_{30}$	*no ~ *ne	*e-
63	**one**$_{21}$	*syo-	*ükːV *V-
65	**rain**$_{39}$	*xeyu- *suwə- (*suH-)	*sʸaða-
78	**smoke**$_{36}$	*dʰuxʷ- (*dʰuh$_2$-)	*sawi **kačV- ~ **kɨčV- **siksV *čüŋV **pusʸV *künti
80	**star**$_{40}$	*xəster- (*h$_2$ster-)	*kuɲsʸV
81	**stone**$_{9}$	**Vsu- *gʷrV-w-on (*gʷreH-won- ~ *gʷr̥H-won-) *akʸ-moːn ~ *xakʸ-moːn (*h$_2$ekʸ-moːn)	*kiwi *pVyi
82	**sun**$_{35}$	*saːw-el- (*seh$_2$w-el-)	*kaya **nVpːV *kečä *päywä
84	**tail**$_{26}$	*puk- **sis- ~ **ses-	*ponči
87	**thou**$_{5}$	*ti ~ *tuː-	*ti-n
88	**tongue**$_{8}$	*dn̥gʰʸwaː ~ *gʰʸn̥dwaː **lal- ~ **lol-	*käli *ɲälmä
89	**tooth**$_{22}$	*ədont- ~ *dont- (*h$_1$dont-) *gʸombʰ- **kVg-	*piŋi **timä ~ **simä
90	**tree**$_{37}$	*doru	*po
91	**two**$_{2}$	*dwo-	*kitːä
94	**water**$_{28}$	*wed-	*weti
95	**we**$_{1}$	*wey-s *n̥s-	*me
96	**what**$_{12}$	*kʷi-	*mi ~ *mɨ
98	**who**$_{6}$	*kʷi-	*ku-

4. Results

Compared to all the previously conducted probabilistic studies on the IE-Uralic connection (Ringe 1998; Oswalt 1998; Kessler & Lehtonen 2006; Kessler 2007, reviewed in §1 above), we believe that our test comes closest to modeling real comparative-historical research, at least as far as criteria for what constitutes an etymological lexical match between two languages are concerned. Indeed:

1) historical linguists implicitly understand that cross-linguistically, the most common root shape is *CVC(V)* (where *C* may be a zero), both consonants of which should correspond to a *CVC(V)* root in the compared language (for instance, linguists do not confine themselves to initial consonants, as some of the aforementioned testers do);
2) although exceptions are common and almost inevitable, the bulk of assumed phonetic shifts should be typologically trivial, i.e., the shifts should happen within the limits of phonetically justified consonant classes;[17]
3) "step-by-step" reconstruction, when a proto-language is reconstructed sequentially on the basis of proto-languages of the previous taxonomic level, is methodologically preferable to direct comparison on the basis of all attested languages at once.[18]

Our statistical results are as follows. Within the accepted 50-item wordlist, we have 7 IE-Uralic pairs which should constitute etymological cognates within the framework of the traditional Nostratic theory as originally outlined by V. M. Illich-Svitych and V. A. Dybo (Illich-Svitych 1967; 1971–1984).

These pairs are (*CC*-transcoding is added in square brackets):

[17] The assumption of a great number of unusual phonetic shift leads to regrettable results, cf., e.g., the critical overview of an IE-Basque hypothesis in Kassian 2013a.

[18] See, for instance, criticism of several such "wholesale" attempts at the reconstruction of African macro-families in G. Starostin 2013.

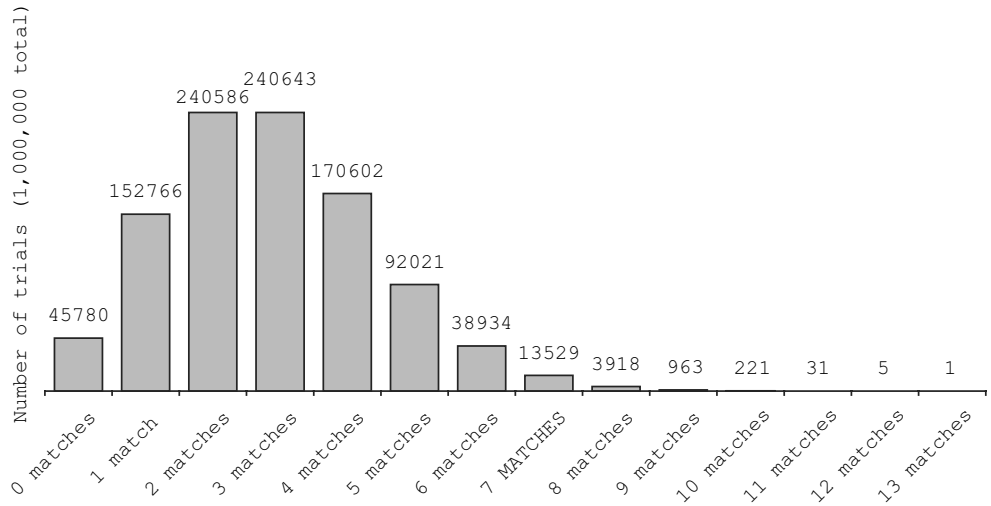
1) IE **kʸlew-* [*KL*] — Uralic **kuwli* [*KL*] 'to hear';
2) IE **me-* [*MH*] — Uralic **mi-n* [*MH*] 'I';
3) IE **nomn̥* [*NM*] — Uralic **nimi* [*NM*] 'name';
4) IE **ti* [*TH*] — Uralic **ti-n* [*TH*] 'thou';
5) IE **wed-* [*WT*] — Uralic **weti* [*WT*] 'water';
6) IE **kʷi-* [*KH*] — Uralic **ku-* [*KH*] 'who';
7) E **eːgʰʷ-* [*HK*] — Uralic **ɨɣi-* [*HK*] 'to drink'.[19]

These and only these pairs are treated as positive matches by our formal algorithm, regardless of whether we use the basic *GLD* consonant classes (Table 1) or narrower classes (Table 2). In other words, our algorithm is adequate enough so as not to detect any parasitic pairs (i.e., pairs that are not connected by regular sound correspondences, accepted in traditional Nostratic studies). It is worth noting that, despite the fact that our Uralic list contains many forms which were reconstructed on the basis of one branch only (in cases where no definite Proto-Uralic reconstruction is possible, see §2.1 above), six out of seven resulting pairs contain strictly Proto-Uralic forms, that is, words present both in Finno-Ugric and Samoyed branches. The remaining form, **kuwli* 'to hear', is at least Proto-Finno-Ugric (attested both in Finno-Permic and Ugric).

Proceeding from the basic *GLD* consonant classes (Table 1), the results of the permutation test are presented in Figure 4.

[19] The seventh pair is not present in Illich-Svitych's (1971–1984) dictionary, since the traditional reconstruction of the Uralic root for 'to drink' is **yuɣe-*, and Uralic *y-* cannot regularly correspond to IE zero. However, recent advances in Uralic etymology now allow us to change the Uralic reconstruction to either **yɨɣi-* or **ɨɣi-*.

Figure 4. IE-Uralic permutation comparison: basic GLD consonant classes



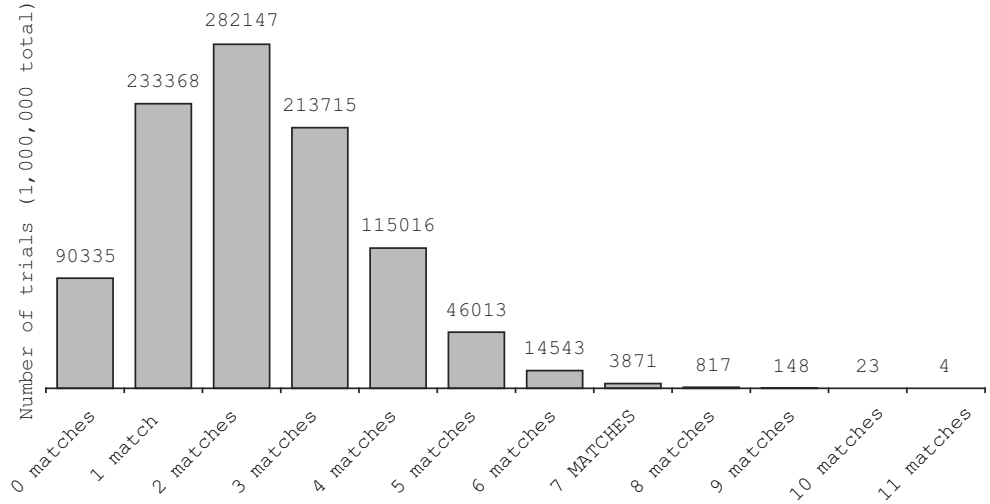
The most statistically common values are 1 match, 2 matches, 3 matches and 4 matches — their probability *P* is 0.152766, 0.240586, 0.240643 and 0.170602 respectively, i.e., *ca.* 15.3%, 24.1%, 24.1% and 17.1% respectively.

The total number of trials with 7 or more matches is 13529 + 3918 + 963 + 221 + 31 + 5 + 1 = 18668. This means that the probability *P* of getting at least seven matches (as we have in the case of the original IE-Uralic list) is 0.018668, i.e., *ca.* 1.9%.

The most frequently accepted level of statistical significance is 5% (it means that the null hypothesis should be rejected if *P*-value is less than 0.05); another popular significance level, used for more precise calculations, is 1% ($P = 0.01$). The probability of IE-Uralic matches (1.9%) is lower than the 5% level, although still slightly higher than the 1% level.

Turning now to calculations based on the more refined subdivision into consonant classes (Table 2), the results of the permutation test are as follows: Figure 5.

Figure 5. IE-Uralic permutation comparison: more precise consonant classes



The total number of trials with 7 or more matches is 3871 + 817 + 148 + 23 + 4 = 4863. This means that the probability *P* of getting at least seven matches (as we have in the case of the original IE-Uralic list) is 0.004863, i.e., *ca.* 0.5%. In this case, the probability of the IE-Uralic matches (0.5%) is lower than the strong 1% level.

The fact that in both cases the obtained probabilistic values (1.9% & 0.5%) are smaller than the common level of statistical significance (5% or even 1%) does not mean that the IE-Uralic matches *cannot* be chance coincidences.[20] And, vice versa, if the IE-Uralic value would be higher than 5%, it would not prove that the observed IE-Uralic matches are necessarily chance coincidences. Nevertheless, general methodology suggests that in the case of such low probability of the IE-Uralic matches it is recommendable to search for a more appropriate explanation

[20] See Kassian 2014 for such unique pairs of unrelated languages as Modern English — Ari (8 *CC*-matches, $p = 0.044\%$) and Abidji — Maidu (7 *CC*-matches, $p = 3.61\%$). In practice, these results are almost immediately ruled out as indicative of relationship due to (a) general historic implausibility (cross-continental ties of this sort and at an obviously deep time length between Europe/Africa and Africa/North America are excluded) and (b) historical linguistic context: all of these languages find closer relatives within their immediate families (Germanic, Omotic, Kwa, Maiduan), and an extended series of pairwise comparisons between other languages of these families will clearly demonstrate that the English/Ari and Abidji/Maidu matches are accidental.

than chance coincidence. The typology of language contacts strongly speaks against a contact-based explanation (see §5). In such a situation, the only solution is the hypothesis of an IE-Uralic genetic relationship, the way it is assumed within the framework of the Nostratic theory.

5. Excursus: Ancient lexical borrowings between Indo-European and Uralic

Although a detailed discussion of Indo-European–Uralic lexical contacts is beyond the scope of the present paper and the previous literature on this topic is enormous, it is useful to briefly outline the main arguments against contact-based explanation of the observed Swadesh matches. It is well known that, among lexical items, cultural vocabulary is always borrowed first, whereas basic vocabulary is generally more resistant to borrowing (Thomason & Kaufman 1988: 74 ff.; Thomason 2001: 70 f.); more precisely, this rule has been derived from the analysis of all the situations in which we have some knowledge of the sociolinguistic history of the specific peoples and languages. However, the etymological connections between IE and Uralic, as they are presented in the original Nostratic hypothesis (Illich-Svitych 1967; 1971–1984), demonstrate a near-complete absence of cultural words: the majority of IE-Uralic etymologies belong to basic vocabulary; see especially Helimski 2000/2001, where IE-Uralic lexical matches are discussed under this theoretical approach. Below we partially repeat Helimski's reasonings.

Two previously published lists of presumed Proto-Indo-European loanwords in Proto-Uralic could be examined: Rédei 1986: 40–43; Koivulehto 2001: 236–238. Taken together, these lists comprise 10 words: 'to give, sell', 'to wash', 'name', 'sinew', 'to bring, give', 'water', 'a kind of metal', 'to fear', 'to plait, to spin', 'to bore'. As noted by Helimski (2000/2001: 498), the meaning 'to sell' in Uralic is obviously derived from 'to give', while notions of 'spinning' and 'boring' "ceased to be technical novelties long before the PIE and PU epoch". Strictly speaking, only Proto-Uralic **wäsʸkä* 'a kind of metal' can be said to belong to cultural vocabulary. This word, however, has straightforward parallels only in Tocharian (Toch. A *wəs*, Toch. B *yasa*

'gold'), so it could be either an early Tocharian loan into separate branches of Uralic (the Uralic word has somewhat irregular vocalic correspondences), or even a loan from Samoyed into Tocharian. Note also that the Rédei / Koivulehto list does not include pronouns and a number of additional parallels in basic vocabulary, regarded as cognates by proponents of the Indo-Uralic hypothesis, e.g., Uralic **kupsa-* 'to extinguish' — IE **g^{w}es-* 'to extinguish' or Uralic **käliw* 'in-law' — IE **g^{y}l̥(x)wos* 'husband's sister', further see IE-Uralic etymologies elaborated by Illich-Svitych (1967; 1971–1984).

The distribution of our seven IE-Uralic matches within the 50-item wordlist is also striking. Four of them fall within the first ten most stable items (as listed in Table 3): 'I'$_{3}$, 'thou'$_{5}$, 'who'$_{6}$, 'name'$_{10}$, with only three belonging to the less stable sections: 'to drink'$_{15}$, 'water'$_{28}$, 'to hear'$_{45}$. Such a distribution within the 50-item wordlist can only be explained by a hypothesis of genetic relationship, since in the case of lexical borrowings, one should expect IE-Uralic matches to be found among the least stable Swadesh items.

Despite the above, there are indeed a lot of lexical items of IE origin which penetrated into Uralic languages in more recent epochs, i.e., after the splits of Proto-Indo-European and Proto-Uralic. The most important loanword strata are Indo-Iranian loans in Finno-Ugric languages (including borrowings in Proto-Finno-Ugric), Iranian loans in Finno-Ugric languages, Baltic loans in Western Finno-Ugric languages (Balto-Finnic, Saami and Mordvinic), Germanic loans in Balto-Finnic and Saami. All these cases confirm the rule that basic vocabulary is borrowed only along with a sufficient amount of non-basic (cultural) vocabulary. As an example we may list the Indo-Iranian loans in Proto-Finno-Ugric, Proto-Finno-Permic and Proto-Finno-Volgaic (Rédei 1986: 43–64). Twenty out of 57 words belong to basic vocabulary in the broadest sense: 'to drive', 'to scatter, throw, dig', 'part, to divide', 'orphan; widow', 'horn', 'to do, make', 'to lead', 'to take, carry', 'totally, exactly', 'side', 'to dig', 'dead', 'man', 'to fasten, tie', 'bottom', 'river, stream', 'shadow', 'salt', '(younger) sister', 'sky, heaven'. The remaining 37 loans belong to non-basic vocabulary (cultural vocabulary and words for flora and fauna):

'price', 'chief, prince', 'hut, house', 'bee', 'honey', 'awl', 'fox', 'thousand', 'hundred', 'ware, merchandise', 'shaft (in horse harness)', 'shoot of grass', 'cereal', 'spindle', 'fish-hook', 'hog', 'slave', 'elm, ash', 'god', 'luck', 'pig', 'rope', 'bridge', 'goat', 'chaff', 'grass, hay', 'sickle', 'beestings', 'ox, bull', 'udder', 'handle', 'wolf', 'male; foal', 'calf', 'axe, hammer', 'cloth, fabric', 'poison'. These loanwords reflect a profound Indo-Iranian influence on Finno-Ugric peoples, especially in the areas of agriculture and cattle-breeding. In the present paper, we discuss this issue sub #41 'horn' and #78 'smoke'.

6. Appendix: Linguistic notes on individual forms

2. 'ashes'. IE: **xaːs-* (**h₂eHs-*) is the only candidate from the topological point of view. It is attested in Anatolian and in a number of Narrow IE groups (Indic, Iranian, Armenian, Germanic). Derived from the verb for 'to dry', but such a derivation must be reconstructed already for the Proto-IE level.

Uralic: the only word for 'ashes' reconstructible for any level above that of individual groups is Ugric **kaðʸma*. E. Helimski suggested that it can be related to Samoyed **kimä*. Although the sound developments here are irregular (we would expect **kɔymɔ* in Samoyed), the connection can be provisionally accepted. Recently, Ante Aikio has proposed an internal etymology for Proto-Ugric **kaðʸma* 'ashes' < Proto-Uralic **kaðʸa-* 'to leave' plus the nominal abstract suffix *-ma* (Zhivlov 2014: 120). Such an analysis faces semantic difficulties (additional instances for semantic derivation 'to leave' > 'leaving' > 'ashes' are needed), but in principle it does not contradict the idea of the Proto-Uralic antiquity of this nominal stem.

6. 'bird'. IE: **awi- ~ *xawi- ~ *wi-* (**h₂ewi- ~ *h₂wei-*) is the best Narrow IE candidate from the topological point of view (attested in Indic, Armenian, Italic). Unfortunately, reliable terms for 'bird' are not attested in Anatolian and Tocharian.

Uralic: every group has its own word for 'bird'. In Samoyed, Ob-Ugric and Balto-Finnic (in Livonian), the

meaning 'bird' is expressed by the general word for 'animal' (**sʸarma, *woːya* and **lintu* respectively), usually in conjunction with such epithets as 'flying' or 'winged'. This semantic pattern may date back to Proto-Uralic, and the extreme instability of the words for 'bird' and 'animal' in Finno-Ugric is probably caused by the breakdown of this pattern under the influence of Indo-European languages.[21] Ob-Ugric **woːya* 'animal / bird' is apparently related to Komi *vəy* 'wild', although the vowel correspondence is not regular. If this etymology is correct, the Ob-Ugric word can be excluded from the list of candidates, since the development was rather from 'wild' to 'animal' than vice versa. Saami **lonteː* 'bird' goes back to Proto-Finno-Ugric **lunta,* whose reflexes mean 'wildgoose' in Ugric and 'duck' in Mari. Permic 'bird' is derived from 'to fly' (cf. probably the same in Balto-Finnic **lintu*), while the Mari word is a Turkic borrowing. The remaining words all have a formal possibility of being inherited from Proto-Uralic. These are: Samoyed **sɔrmɔ* < ***sʸarma,* Hungarian *mədaːr* < ***mVntV,* Mordvinic **narmə̌nʸ* < ***nVrmV-*.

8. 'black'. IE: There are 3 formally equal candidates. (1) **kr̥s-no-* (Indic, Baltic, Slavic); (2) **melə-n-* (**melh₂-n-*) (Greek, Baltic); (3) **dʰengʷ-* ('black, dark' in Anatolian, 'dark' in Germanic).

Uralic: The Proto-Samoyed word for 'black' is not reconstructible: almost every language has its own root, and most lack reliable etymologies (except for Mator *küːnda,* which seems to go back to 'smoke'). While the Selkup-Kamassian isogloss (Selkup *säːg,* Kamass. *saːgar*) looks promising, it is, unfortunately, probably an old borrowing from Tungus-Manchu. The following Finno-Ugric forms can be listed among the candidates: Balto-Finnic **musta-* < ***musi-,* Saami **čaːpːeː* < ***sʸäpːä,* Mordvinic **ravžə̌* (cognate to Saami **ruopseː* 'red' < Proto-Finno-Ugric **rapša*), Permic **sʸəd* < ***sʸɨnti,* Proto-Ugric **pekːVtːV.* Mari **simə̌,*

[21] It is interesting that Balto-Finnic (Livonian) **lintu* 'beast / bird' can be viewed as a derivative from the verb 'to fly' (**lentä-* in most Balto-Finnic languages, but **lintä-* in Livonian), or may represent the result of secondary contamination with this verb.

cognate to or borrowed from Permic **sim* 'rust', can be excluded from the list.

9. 'blood'. IE: **esx-r̥* (**h_1esh_2-r̥*) is the only candidate. It is attested in Anatolian and in the majority of Narrow IE groups.

Uralic: there are two equal candidates: Finno-Ugric **weri* and Samoyed **kəm* < ***kɨmi*.

10. 'bone'. IE: **xost-* (**$h_2$est-*) is the only candidate from the topological point of view. It is attested in Anatolian and in the majority of Narrow IE groups.

Uralic: **luwi* is the only candidate from the topological point of view. It is attested in Samoyed and in the majority of Finno-Ugric groups.

13. 'fingernail'. IE: phonetically and morphologically, this is a non-trivial case. Technically, we reconstruct several variants: **θmog^h-* ~ **θn̥g^h-* ~ **mog^h-* ~ **nog^h-* ~ **n̥g^h-* (**$h_3$nog^h-* ~ **h_3n̥g^h-*), regardless of whether these originate from a single ablaut proto-stem or not (for **θ* see §3 above). Actually we suppose that the original shapes of the stem were ablaut **θmog^h- / *θm̥g^h-*. The full-grade variant **θmog^h-* is directly attested in Anatolian (Luwian *tam͗uga-* with an epenthesis in the initial consonant cluster and occasional labialization *a* > *u* near *m*) and Tocharian **mek*(-)*wa͗*. The zero-grade variant **θm̥g^h-* became **θn̥g^h-* elsewhere, since *n* and *m* are neutralized before velars; **θn̥g^h-* is attested in Anatolian (Hittite *sank*(-)*uwai-*) and in some Narrow IE languages (Italic, Celtic). The *schwebeablaut* variant **nog^h-*, attested in many Narrow IE groups (Greek, Armenian, Indic, Baltic, Slavic, Germanic), is a back-formation from **n̥g^h-* (< **θn̥g^h-*).

Uralic: there are two equal candidates: Finno-Ugric **künči* and Samoyed **kəðɔ* < ***kuta*.

17. 'to die'. IE: **wel-* is the best candidate from the topological point of view. It is attested in Anatolian (Luwian) and Tocharian. In Proto-Narrow IE, **wel-* was euphemistically superseded by **mer-* 'to disappear' (as proven by the Hittite data).

Uralic: **kali-* is the only candidate. It is retained in

Samoyed and the majority of the Finno-Ugric groups.

18. 'dog'. IE: **kʸwon-* is the only candidate. It is attested in Anatolian and in the majority of Narrow IE groups.

Uralic: There are three equal candidates: Samoyed **wən* < ***wɨni*, Ugric **ämpV* and Finno-Permic **penä*.[22]

19. 'to drink'. IE: **e͡ːgʰʷ-* (**h₁egʰʷ-*) is the best candidate from the topological point of view. It is attested in Anatolian and Tocharian, whereas in Proto-Narrow IE, **e͡ːgʰʷ-* was superseded by **poːy- / *piː-*.

Uralic: As was shown by Ante Aikio (2002: 38–40), the Proto-Samoyed verb **ər-* 'to drink' is actually cognate with Proto-Finno-Ugric **yɨɣi-* 'to drink'. The Samoyed stem should be analyzed as **ə-r-* with the augmentative suffix **-r-*, cf. another Proto-Samoyed derivative from this root: **ə-kə̆l-* 'to drink avidly'. The resulting Proto-Samoyed root **ə-* 'to drink' can go back to Uralic **ɨɣi-* without initial **y-*. Although Aikio assumes a sporadic loss of the initial glide in Proto-Samoyed, strictly speaking, we cannot decide which of the two variants — with or without **y-* — is primary for Proto-Uralic. We include both forms, **yɨɣi-* and **ɨɣi-*, in the list.

20. 'dry'. IE: **saws-* is the best candidate for the status of the Proto-Narrow IE term for 'dry' and eventually of the Proto-IE one as well, since in Anatolian (Hittite) and Tocharian, expressions for 'dry' look like recent introductions: the synchronic participle from the verb *xad-* 'to dry' in Hittite and the synchronic forms of the verb *aːs-* 'dry out, dry up' in Tocharian.

Uralic: **kosʸka* is the best candidate from the topological point of view. This word is attested in Samoyed and in some Finno-Ugric groups (Saami, Mordvinic) with derivatives in Mari and Permic.

21. 'ear'. IE: there are two equal candidates: ***stom-* (Anatolian) and **ows-* ~ **xows-* (**h₂ews-* ~ **h₂ows-*) (Narrow IE). In Tocharian, 'ear' is a transparent deverbative from

[22] Note that Samoyed **wən* could theoretically represent a loanword from an IE language such as, e.g., Proto-Tocharian.

'to hear'.

Uralic: there are two equal candidates: Finno-Ugric **pelyä* and Samoyed **kɔw* < ***kawi*. The Proto-Uralic antiquity of the Samoyed word for 'ear' is, however, doubtful: other Proto-Uralic stems of the shape **(C)Vwi* yield **(C)V* in Samoyed.

23. 'to eat'. IE: **ed-* (**$h_1$ed-*) is the only candidate from the topological point of view. It is attested in Anatolian and in the majority of Narrow IE groups.

Uralic: there are two candidates: Finno-Ugric **sewi-* (retained in Baltic-Finnic, Permic, Ob-Ugrian, Hungarian) and Samoyed **ə̆m-*. The Samoyed word goes back to Proto-Uralic **imi-*, whose Finno-Ugric reflexes mean 'to suck'. Since we cannot exclude the possibility of a semantic development 'to eat' > 'to suck' in Finno-Ugric, we list both candidates.

24. 'egg'. IE: **o:wyo- ~ *xo:wyo-* (**h_2o:wyo-*) is the only Proto-Narrow IE candidate and, eventually, the only one for Proto-IE (unfortunately, the Anatolian and Tocharian terms for 'egg' remain unknown). Reconstruction of medial glides is unclear: we postulate the "cumulative" cluster **wy*, but the transcriptions with single **w* or **y* would yield the same consonantal skeleton *HH* as **o:wyo-*.

Uralic: **muna* is the only candidate. It is attested in Samoyed and in the majority of Finno-Ugric groups.

25. 'eye'. IE: we reconstruct two variants, **θok^w- ~ *ok^w-* (**$h_3$ek^w-*), where **ok^w-* is the indisputable Proto-Narrow IE candidate, whereas **θok^w-* is the Proto-IE form if we include the Anatolian data (Hittite *saguwa-*, Luwian *tawi-*) in the comparison; for **θ* see §3 above.

Uralic: **sʸilmä* is the only candidate, attested in all the groups.

28. 'fire'. IE: **pex-wr̥* (**peh$_2$-wr̥*) is the best candidate from the topological point of view. It is retained in Anatolian, Tocharian and many Narrow IE groups (Greek, Armenian, Germanic, Italic). The second candidate is **n̥g-n-i-* (Indic, Baltic, Slavic, Italic), which competes with **pex-wer-* in a "criss-crossed" situation from the topological point of view. Cf. especially the Italic group, where the first stem is used to denote 'fire' in Umbrian and the second one — in

Latin. Despite the fact that **n̥g-n-i-* looks much weaker, we treat both as synonyms.

Uralic: **tuli* is the best candidate. It is preserved in Samoyed and Finno-Permic, but is replaced by **tüɣVtV* in Ugric.

31. 'foot'. IE: **pod-* is the only candidate from the topological point of view. It is attested in Anatolian and in the majority of Narrow IE groups.

Uralic: an unstable word. **yalka* seems to be the best candidate for the status of the Proto-Finno-Ugric term, whereas Samoyed has **ɔy* < ***ayi* or ***ali*.

36. 'hair'. IE: an unstable word. First, we postulate two equal roots, attested in the Anatolian group: ***ted-* (Hittite *ted*(-)*ana-*) and ***top-* (Luwian *tapː*(-)*ani-*), both without external etymology. As for Narrow IE, there are two main candidates in the topologically "criss-crossed" configuration, which we have to treat as synonyms. (1) The first one is **dekʸ-* / **dokʸ-*, preserved as Tocharian A *sʸaːku* '(head)hair' and Gothic *tag-l* '(head)hair' (in other Germanic languages, **tag-la-* means 'hair of tail', 'tail', 'rope's end', but if Gothic is the first outlier, **tag-la-* can be posited as the Proto-Germanic term for 'hair'). Further to Old Indic *dasʸaː-* 'fringe of a garment, loose ends of any piece of cloth'. Both consonantal (*sʸ*) and vocalic (*aː*) reflexes in the Tocharian form are not fully regular, but at least the consonant outcome is the same as in the Tocharian word for '10'. (2) The second Narrow IE candidate is **wel-* with the suffixal extensions *-so-* (Iranian, Slavic) or *-to-* (Celtic). In Indic and Greek, the original term has apparently been replaced by loanwords.

Uralic: **i̯pti* is the best candidate. It is attested in Samoyed and in many Finno-Ugric groups (Saami, Mari, Ob-Ugric).

37. 'hand'. IE: **gʰʸes-r̥* is the best candidate from the topological point of view. It is attested in Anatolian and in many Narrow IE groups.

Uralic: there are two equal candidates: Finno-Ugric **käti* (retained in all the groups) and Samoyed **utɔ* < ***VtV* (Proto-Uralic **uta* would have given **ŏtɔ* in Samoyed).

38. 'head'. IE: a difficult case with several competing roots. (1) The first candidate is **kʸer- / *kʸr̥-*, which is attested in the meaning 'head' in some Narrow IE groups with or without suffixal extensions (Indic, Iranian, Greek, Albanian), as well as residually preserved in Anatolian (the Hittite adverb *kit-kar* 'at the head of'). The root **kʸer-* is also attested in the stems for 'horn' among IE languages, but the latter meaning is secondary, since only the semantic derivation 'head' > 'horn' is typologically normal, not vice versa. (2) The second candidate is **gʰebʰ-Vl-*, which means 'head' in Tocharian and Ancient Greek (the second Greek root for this basic meaning!), and 'gable' in Germanic. (3) Finally, the third candidate is Anatolian **xars-* 'head', attested as Hittite heteroclitic *xars-ar* 'head' and Luwian suffixed *xarmaxi-* 'head' (< **xars-maxi-*?). Theoretically, it is possible to analyse the Anatolian form as **xar-s-* with additional suffixation and derive it from IE **kʸr̥-s-* with irregular fricativization *k* > *x*. Formally, we fill the IE slot with three synonyms: **kʸer-, *gʰebʰ-Vl-* and ***xors-*.

Uralic: **oywa* is the best candidate from the topological point of view. It is attested in Samoyed and in some Finno-Ugric groups (Saami, Mari).

39. 'hear'. IE: **kʸlew-* is the best candidate for the status of the Proto-Narrow IE term (it is attested in the majority of Narrow IE groups) and eventually of the Proto-IE one, since in Anatolian, the verbs for 'to hear' look like denominative stems from the words for 'ear'.

Uralic: there are two equal candidates: Finno-Ugric **kuwli-* (retained in all the groups) and Samoyed **yüntə̆-* < ***yünti-*. The idea that **kuwli-* is derived from the root ***kawi* 'ear' (q.v.), reflected in Samoyed **kɔw*, remains speculative, since the development **aw* > **uw* is not confirmed by additional examples. Moreover, the direct semantic evolution 'ear' > 'to hear' seems to be typologically abnormal; a far more common derivation would be 'ear' > 'to listen', which would presume an additional hypothetical stage in such a Proto-Uralic scenario: 'ear' > 'to listen' > 'to hear'.

40. 'heart'. IE: **kʸerd-* is the only candidate. It is attested in all the groups except for Tocharian.

Uralic: **sʸäðʸä* is the only candidate. It is attested in all the groups.

41. 'horn'. IE: in the majority of groups (including Anatolian and Tocharian), terms for 'horn' represent various suffixal formations from the root **kʸer- / *kʸr̥-* 'head' q.v. The most archaic suffixal pattern seems to be **kʸr-ew-r̥* (Hittite *karawar*, Tocharian **kror-*), later stems are **kʸer-w-*, **kʸr̥-no-*, **kʸer-Vs-*.

Uralic: Samoyed **amtə̆* < ***amti* is the most reliable candidate. Almost everywhere in Finno-Ugric, inherited terms were superseded by an Indo-Iranian loanword.

42. 'I'. IE: the suppletive paradigm **egʸ-* [nom.] / **(ə)me-* [obl.] (**h₁egʸH- / *h₁me-*) can be reconstructed with safety.

Uralic: **mi-n* is retained in all the groups. Final **-n* is treated here as a suffix, since it is absent in the evidently related pronoun **me* 'we'.

43. 'to kill'. IE: **gʰʷen-* is the best candidate from the topological point of view. It is attested in Anatolian and in some Narrow IE groups (Indic, Iranian), having been superseded by various unrelated verbs in the rest of Narrow IE. Since **gʰʷen-* is more widely attested with the meaning 'to hit' *vel sim.* (e.g., Greek and so on), and it is the semantic shift 'to hit' > 'to kill' that is natural, but not vice versa, we should reconstruct **gʰʷen-* with the polysemy 'to kill / to hit' already for the Proto-IE level.

Uralic: an unstable word. In Samoyed, 'to kill' is expressed by the causative from 'to die' (**kɔ-tɔ-*), a recent introduction. In Finno-Ugric, the best candidate from the topological point of view is **wiði-* (Permic, Ob-Ugrian, Hungarian; the Saami reflex means 'to beat').

46. 'leaf'. IE: ***porst-* is formally reconstructible on the basis of the isolated Hittite form *parstu-*. The best Narrow IE candidate is **bʰol-* with different suffixal extensions (attested in Tocharian, Greek, Germanic, Italic).

Uralic: reflexes of **lVpV-* (vowel correspondences are irregular) are preserved in Samoyed and Ugric. Finno-Volgaic has replaced this word with the Balto-Slavic borrowing **lešti*.

48. 'liver'. IE: there are four main candidates: ***les-* (Anatolian: Hittite), ***pVn-* (Anatolian: Palaic), **ud-tr̥-yo-* (Tocharian), **yekʷ-r̥* (Narrow IE minus Tocharian). Out of these, Tocharian **ud-tr̥-yo-* is morphologically secondary and therefore has a lesser chance to represent the original term for 'liver', but **les-*, **pVn-* and **yekʷ-r̥* are formally equal variants.

Uralic: **mɨksa* is the only candidate. It is attested in all the groups.

53. 'meat'. IE: **mems-* is the best candidate. It is attested in the majority of the Narrow IE groups including Tocharian. The Anatolian term for 'meat' has not been documented so far.

Uralic: an unstable word. 3 topologically equal candidates can be listed: **siwVlV* (Mordvinic, Mari, Permic), **ɲəɣVðʸ* < ***ɲVkVðʸV* (Ob-Ugric), **ɔyɔ* < ***aya* ~ ***oya* ~ ***aðʸa* ~ ***oðʸa* (Samoyed).

54. 'moon'. IE: there are two topologically equal candidates: ***or-* (based on Anatolian **ar-ma-*) and Narrow IE *meːn-oː-t*, obl. *meːn-(e)s-* (**meh₁n̥s-*)[23].

Uralic: **kuŋi* is the best candidate. It is retained in Samoyed and in the majority of the Finno-Ugric groups.

56. 'mouth'. IE: **o(ː)y-es*, obl. *e(ː)y-(e)s-* vel sim. (**HoH-es-*) is the best candidate. It is attested in Anatolian and in some Narrow IE groups (Iranian, Italic, Celtic). Morphological details are, however, unclear: probably the most archaic paradigm is retained in Hittite, whereas Narrow IE languages demonstrate the analogically levelled stem **oːs-*.

Uralic: there are two equal candidates: Finno-Ugric **sʸuwi* (attested in Baltic-Finnic, Mansi and Hungarian) and Samoyed **aŋ* < ***aŋi*.

57. 'name'. IE: **nom-n̥* ~ **lom-n̥* ~ **ənom-n̥* ~ **əlom-n̥*

[23] The laryngealistic reconstruction (supported by the Indo-Iranian reflex **maːs-* < **maas-* with vocalisation of syllabic **n̥*) means that **meh₁-n̥s-* is most probably derived from the verb **meh₁-* 'to measure'. If this is correct, Anatolian **or-mo-* becomes the most likely candidate. Actually, however, semantic derivation 'to measure' > 'moon' is atypical cross-linguistically.

(*h_1nom-$n̥$ ~ *$h_1n̥m$-en-), this stem is retained in all the groups, but technically we reconstruct several phonetic shapes (the variants with the initial schwa are reconstructed because of the Luwian and Greek cognates). We take issue with the frequently assumed reconstruction of the IE word for 'name' with an internal laryngeal (*h_1neh_3-*men*- or *h_3neh_3-*men*-). Phonological and morphological arguments against this reconstruction are summed up by K. Stüber (1998: 53–59). Kloekhorst (2008b: 518) has suggested that *h_3neh_3-*men*- 'name' is derived from the verbal root *h_3neh_3- 'to call (by name)', reflected in Hittite *xann(a)*- 'to sue; to judge' (< *'to call to court') and Ancient Greek *ónomai* 'to scold, blame, insult' (< *'to call names'). While the development 'to call by name' > 'name' is typologically plausible, the reconstructed meaning 'to call by name' is itself rather based on comparison with the noun 'name' than on the attested meanings of the Hittite and Greek verbs. The semantic reconstruction based on the latter would rather be 'to blame, to reproach'. Note also that Kloekhorst's hypothesis implies that the negative semantics of the Hittite and Greek verbs developed not only independently, but along different lines.

Uralic: **nimi* is the only candidate. It is retained in all groups. In Mordvinic and Mari, dissimilative forms with initial *l*- are observed, but due to topological reasons, these cannot be projected onto the proto-level (as opposed to the IE situation).

59. 'new'. IE: **new*- is the only candidate. It is attested in all the groups except for Albanian.

Uralic: The Proto-Samoyed word is not reconstructible (every language has its own root), therefore, we are left with Finno-Ugric **wuðʸi* (attested in all Finno-Ugric groups except for Ob-Ugrian).

60. 'night'. IE: **ksep*- / **ksp*- is the best candidate. It can be reconstructed on the basis of Anatolian data (Hittite *sp-ant*-); in Proto-Narrow IE, **ksep*- was superseded by *nog^{hw}-*t*-, which represents a deverbative formation from the stem 'to become evening' (as proven by Hittite). It is interesting that both **ksep*- and *nog^{hw}-*t*- demonstrate a "zig-zagging" development in Indo-Iranian: first, in Proto-Indo-Iranian, **ksep*- secondarily acquired the basic meaning

'night', whereas **nog^{hw}-t-* was retained in the adverbial function 'at night', but **nog^{hw}-t-* nevertheless managed to become the generic term for 'night' in Wakhi.

Uralic: there are two equal candidates: Finno-Ugric **eyi* (attested in all groups except for Mari) and Samoyed **pi* < ***piyV* ~ ***piγV*.

61. 'nose'. IE: there are 3 topologically equal candidates. (1) ***tid-*, which can be formally reconstructed on the basis of Anatolian (Hittite) data. (2) ***mVl-*, which can be formally reconstructed on the basis of the Tocharian data. And finally (3) **nas-*, which is attested in a number of Narrow IE groups (Indic, Iranian, Baltic, Slavic, Germanic, Italic).

Uralic: there are three equal candidates: Samoyed **pɨyɔ* < ***pɨya* ~ ***pɨðʸa*, Ob-Ugric **ɲulV* < ***ɲulV* ~ ***ɲuðV*, and Finno-Permic **näri*.

62. 'not'. IE: proclitical **no* ~ **ne* is the best candidate. It is attested as the basic verbal negation of assertion in Anatolian and in the majority of the Narrow IE groups.

Uralic: the negative verb **e-* is the best candidate. It is attested as the basic verbal negation of assertion in Samoyed and in all Finno-Ugric groups except for Hungarian.

63. 'one'. IE: **syo-* is the best candidate from the topological point of view. It is attested in Anatolian (Hittite) and Tocharian, having been superseded by forms of the shapes **oy-no-, *oy-ko-, *oy-wo-* in other groups.

Uralic: there are two candidates: Samoyed **o-* < ***V-* and Finno-Ugric **ükːV*, modified to **ükti* in Finno-Volgaic (judging by Mari data, **ük-ti* was originally a non-attributive form, opposed to attributive **ükːV*).

65. 'rain'. IE: there are two main candidates. The more preferable one is the nominal stem **xeyu-*, reconstructed on the basis of Anatolian. The weaker one is the verb **suwə-* (**suH-*) 'to rain', from which various unrelated nominal forms denoting 'rain' are derived in Tocharian, Greek and Albanian. If **suwə-* is indeed cognate with Hittite *suxːa-* 'to pour (out), scatter', the Anatolian (Hittite) meaning should be considered original, whereas the meaning 'to rain' is a Proto-Narrow IE innovation.

Uralic: Proto-Uralic verb **sʸaðа-* 'to rain' is reflected in Samoyed **sɔrɔ-* 'to rain' (deverbal noun **sɔr-ö* 'rain') and Balto-Finnic **sata-* 'to rain'. Although the word for 'rain' was highly unstable in Finno-Ugric, we can safely suppose that in Proto-Uralic the noun 'rain' was derived from the verb **sʸaðа-*.

78. 'smoke'. IE: **dʰuxʷ-* (**dʰuh₂-*) is the best candidate. It is originally a verbal root ('to smoke'), from which various suffixed stems denoting 'smoke' are attested in Anatolian and many Narrow IE groups (Indic, Iranian, Baltic, Slavic, Italic).

Uralic: there are six candidates: Finno-Saamic **sawi*, Mordvinic **kačV-* < ***kačV-* ~ ***kɨčV-*, Mari **sɨ̆ks* < ***siksV*, Permic **čɯŋ* < Finno-Ugric **čüŋV*, Ob-Ugric **posʸV* < ***pusʸV* (despite Rédei 1988–1991: 879, Hungarian *füšt* 'smoke' is an Iranian loanword, unrelated to Ob-Ugric **posʸV*, Helimski 2002: 109), and Samoyed **küntə̆* < Finno-Ugric **künti* (Finno-Ugric reflexes of this root mean 'fog, mist'). The latter word is probably the best candidate, since it was certainly present in Proto-Uralic, whereas the distribution of most of the other candidates is quite narrow.

80. 'star'. IE: **xəster-* (**h₂ster-*) is the best candidate. It is attested in all the groups except for Albanian, Baltic and Slavic.

Uralic: **kuŋsʸV* is the best candidate from the topological point of view. It is attested in Samoyed and some Finno-Ugric groups (Permic, Ob-Ugric).

81. 'stone'. IE: an unstable word with several competing proto-forms. (1) The first formal candidate is the isolated term ***Vsu-*, attested in Anatolian as Luwian *asu-* ⟨á-su-⟩ 'stone' (the Hittite term for 'stone' remains unknown; *agu-* rather means 'seashell'). (2) The second candidate is the suffixal formation **gʷrV-w-on* (**gʷreH-won-* ~ **gʷr̥H-won-*), preserved as Tocharian B *kərweŋe* 'stone' (perhaps < **gʷrV-w-on-en*, extended with additional *-en* — a productive pattern in Tocharian); in Narrow IE (Celtic, Indic), **gʷrV-w-on* is attested as *terminus technicus* 'millstone; stone for pressing the soma' (note that the semantic development 'stone' > 'millstone' is much more natural than vice

versa).[24] (3) The third candidate is Narrow IE **akʸ-moːn* ~ **xakʸ-moːn* (**h₂ekʸ-moːn*) ('stone' in Indic, Iranian, Baltic, Slavic; 'anvil' in Greek); it is the weakest one, since the suffix *-moːn* points to a deverbal derivation, although the meaning of presumed **akʸ-* is unclear ('to be sharp'?).

Uralic: there are two equal candidates: Finno-Ugric **kiwi* (attested in the majority of groups except for Saami and Permic) and Samoyed **pə̆y* ~ **pɔy* < **pVyi*.

82. 'sun'. IE: **saːw-el-* (**seh₂w-el-*) (with ablaut variants) is the best candidate. It is attested in the majority of the Narrow IE groups except for Tocharian (where the deverbative from 'to heat' *vel sim.* is used), Armenian and Albanian. In Anatolian (Hittite), only the Hattic loanword is attested.

Uralic: an unstable word. Proto-Samoyed has **kɔyɔ* 'sun' (opposed to **yalä* 'day, light' that secondarily shifts its meaning to 'sun' in some Samoyed languages). This word may be related to the Finno-Ugric verb **kaya-* 'to become visible, appear'. The Ob-Ugric word (Mansi **kaːtə̆l* 'day, sun', Khanty **kɨtə̆l* 'id.') is a deverbal noun from the verb 'to dawn' (Khanty **katə̆l-* 'to dawn').[25] Hungarian has *nɔp* (no etymology, formally may go back to ***nVpːV*). Permic **šɵnd-i* 'sun' is a transparent derivative of **šɵnɨd* 'warm'. Mari and Mordvinic words go back to **kečä*. Traditionally these words are compared to Permic **kʉč* 'ring' and Balto-Finnic **kehä-* 'id.'; the implication is that the semantic shift 'ring' > 'sun' has occurred in Mari and Mordvinic. There are several problems with this etymology. First, we do not know of any other cases in the world's languages when the meaning 'sun' develops from 'ring'. Second, the Permic form is evidently cognate to Ob-Ugric **küčV* 'ring', thus requiring the reconstruction of **ü* in the first syllable, incompatible with the Mari, Mordvinic and even Balto-Finnic forms. Saami and Balto-Finnic have **päywä* < Proto-Uralic **päywä* (its Samoyed reflex means 'warm, heat').

---

[24] Tocharian A *paːrə̃* can hardly mean generic 'stone' (etymologically, it originates from the meaning 'rock, cliff', as may be seen from the Hittite and Old Indic data).

[25] The ablaut *low vowel* > *high vowel* is a regular means of deriving abstract nouns from verbs in Proto-Khanty.

Permic and Ob-Ugric words must be excluded as clearly derivative; **päywä* is the most probable of the four remaining candidates (the semantic shift 'sun' > 'heat' is attested, e.g., in Selkup, where Proto-Samoyed **yalä* 'day, light' shifted its meaning to 'sun', whereas **kɔyɔ*, the old word for 'sun', came to mean 'heat').

84. 'tail'. IE: there are two topologically equal candidates: ***sis-* ~ ***ses-*, suggested by the Anatolian (Hittite) data, and Narrow IE **puk-* (Tocharian, Indic, further to the Germanic word for 'fox'). Generally, 'tail' is a very unstable word in IE, even within individual groups.

Uralic: the only acceptable candidate is Proto-Uralic **ponči*. This word is the basic designation of 'tail' in Mari and Permic; the meaning of its Samoyed reflex **pɔnčə* 'hem, lower edge' has clearly evolved from 'tail' (the meaning 'hem' is also attested in Komi). Forms with the meaning 'tail' in other branches do not have satisfactory etymologies. Note that Samoyed **tɔywɔ* 'tail' cannot be related to Saami **seaypeː* 'id.' for phonological reasons.

87. 'thou'. IE: on the basis of Anatolian, the suppletive paradigm **ti(ː)* [nom.] / **tu(ː)-* [obl.] can be reconstructed (after the split of Anatolian, the nominative form was analogically levelled after the oblique stem).

Uralic: **ti-n* is the only candidate. It is attested in all the groups except for Ob-Ugric, where the unclear form **neŋ* 'thou' is observed instead — perhaps as a result of influence on the part of the verbal 2nd p. sg. ending of the subjective conjugation. Final **-n* in **ti-n* is treated here as a suffix, since it is absent in the evidently related pronoun **te* 'you' (pl.).

88. 'tongue'. IE: there are two topologically equal candidates: ***lal-* ~ ***lol-*, formally suggested by the Anatolian (Hittite) data, and Narrow IE **dn̥g^{hy}waː* ~ **g^{hy}n̥dwaː*. Cognates of the latter in individual groups are very irregular (fanciful phonetic irregularity is not rare for terms for 'tongue' cross-linguistically), note particularly the contamination with the verb 'to lick' in Armenian, Baltic, Italic. Nevertheless, the match between Tocharian, on the one hand, and Old Latin, Germanic and Celtic, on the other, proves that non-lateral **d* (in Celtic **t* due to assimi-

lation) is the original consonant in this stem. It is impossible to determine, however, which of the two variants, **dn̥g^{hy}wa:* (Narrow IE minus Tocharian) or **g^{hy}n̥dwa:* (Tocharian), is primary.

Uralic: **käli* is the best candidate from the topological point of view. It is attested in Samoyed and in some Finno-Ugric groups (Balto-Finnic, Mordvinic, Permic). In the rest of Finno-Ugric, **käli* is retained with the meanings 'word', 'speech' and so on. The second candidate, **ɲälmä,* also goes back to at least Proto-Finno-Ugric, since it is attested with the meaning 'tongue' in Mari, Ob-Ugrian and Hungarian. The difference in meaning between **käli* and **ɲälmä* is hard to establish. We treat both words as synonyms, just like in the case with IE 'fire' q.v.

89. 'tooth'. IE: there are 3 candidates: ***kVg-* (Anatolian), Narrow IE **ədont- ~ *dont-* (**h$_1$dont-*) and **g^{y}omb^{h}-*. The latter two occur topologically in a "criss-crossed" situation; consequently, we accept both as synonyms. The traditional analysis of **(ə) dont-* as a participle of 'to eat' faces serious typological difficulty: we are not aware of such a semantic derivation cross-linguistically.[26] **g^{y}omb^{h}-* means either 'tooth' or 'peg, sharp stick' across the languages, but semantic shift between 'tooth' and 'peg, sharp stick' is possible in both directions.

Uralic: there are two equal candidates: Finno-Ugric **piŋi* (attested in the majority of groups except for Baltic-Finnic and Saami) and Samoyed **timä* < ***timä* or ***simä.*

90. 'tree'. IE: **doru* is the best candidate from the topological point of view. It is attested in Anatolian and some Narrow IE groups (Greek, Albanian, Slavic, Germanic).

Uralic: **po*[27] is the best candidate from the topological

[26] Unless the Pre-Proto-IE meaning of **ed-* was 'to bite' (with the normal shift 'to bite' > 'to eat' in Proto-IE), since the derivation 'to bite' > 'tooth' is typologically possible. However, there is no unequivocal evidence for an original **ed-* 'to bite' (although cf. Schindler 1975).

[27] Usually reconstructed as **puwe* or **puxi.* However, the reflexes point rather to **o* as a first syllable vowel and do not support the reconstruction of a second syllable. Apparently, this is one of the **CV*-stems, extremely rare in Proto-Uralic.

point of view. It is attested in Samoyed and in many Finno-Ugric groups (Baltic-Finnic, Mari, Permic, Hungarian).

91. 'two'. IE: **dwo-* is the only candidate. It is attested in all subgroups.

Uralic: **kitːä*[28] is the only candidate, preserved in all subgroups.

94. 'water'. IE: **wod-r̥ / *ud-(e)n-* is the best candidate from the topological point of view. It is attested in Anatolian and the majority of the Narrow IE groups except for Iranian, Armenian, Italic.

Uralic: **weti* is unquestionably the best candidate, preserved everywhere except Saami and Khanty.

95. 'we'. IE: the suppletive paradigm **wey-s* [nom.] / **n̥s-* [obl.] can be safely reconstructed, primarily on the basis of Anatolian, Indic and Germanic data. The main riddle concerns those forms that show the shape **me-(s)*, which supersede **wey-s* in Armenian, Baltic and Slavic. Their proto-status is problematic: most likely, they have been carried over by analogy either from the verbal 1st p. pl. ending or the 1st p. sg. pronoun or both, so we avoid to reconstruct **me-* as one more stem within this pronominal paradigm.

Uralic: **me* is the only candidate. It is retained in all the groups.

96. 'what?'. IE: **kʷi-d* is the only candidate. It is retained in all the groups.

Uralic: **mi ~ *mɨ* is the only candidate. It is retained in all the groups.

98. 'who?'. IE: **kʷi-s* is the only candidate. It is retained in all the groups.

Uralic: **ku-* is the only candidate. It is retained in all the groups.

---

[28] Usually reconstructed as **kektä* vel sim. Since **-kt-* regularly yields **-ɣt-*, not **-t-* in Ob-Ugric, we prefer the reconstruction **kitːä*, fully compatible with the Samoyed and Hungarian data. The shape of this numeral in other branches (Permic **kʉk*, Finno-Volgaic **kakta*) is heavily influenced by the numeral 'one'. The differences in vocalism are hard to explain; **i* in our reconstruction is based mainly on Samoyed evidence.

Supplementary materials

Proto-IE and Proto-Uralic 50-item wordlists (Tab. 4) in MS Excel-format. Can be downloaded at:

http://JIES.org/DOCS/IE-Ural-wordlist.xls

References

Aikio, Ante
2002 New and old Samoyed etymologies. *Finnisch-Ugrische Forschungen* 57: 9–57.

Anthony, David W. & Don Ringe
2015 The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1: 199–219.

Baxter, William H.
1995 'A stronger affinity … than could have been produced by accident': A probabilistic comparison of Old Chinese and Tibeto-Burman. In: W. S-Y. Wang (ed.). *The Ancestry of the Chinese Language* (Journal of Chinese linguistics. Monograph series 8), 1–39. Berkeley: University of California.
1998 Response to Oswalt and Ringe. In: J. C. Salmons & B. D. Joseph (eds.). *Nostratic: sifting the evidence*, 217–236. Amsterdam: Benjamins.

Baxter, William H. & Alexis Manaster Ramer
1996 Review of: D. A. Ringe. On Calculating the Factor of Chance in Language Comparison. *Diachronica* 13: 371–384.
2000 Beyond lumping and splitting: Probabilistic issues in historical linguistics. In: Colin Renfrew, April McMahon & Larry Trask (eds.). *Time Depth in Historical Linguistics*, 167–188. Cambridge, England: McDonald Institute for Archaeological Research.

Beekes, Robert S. P.
2010 *Etymological Dictionary of Greek*. Leiden / Boston: Brill.

Blažek, Václav
2007 From August Schleicher to Sergei Starostin: On the development of tree-diagram models of the Indo-European languages. *JIES* 35: 82–109.

Brown, Cecil H., Eric W. Holman & Søren Wichmann
2013 Sound correspondences in the world's languages. *Language* 89/1: 4–29.

Burlak, Svetlana & Sergei Starostin
2005 *Sravnitel'no-istoricheskoe yazykoznanie* [*Historical linguistics*]. 2nd ed. Moscow: Academia.

Campbell, Lyle & William J. Poser
2008 *Language Classification: History and Method*. New York: Cambridge University Press.

Collinder, Björn
1934 *Indo-uralisches Sprachgut: die Urverwandtschaft zwischen der indoeuropäischen und der uralischen (finnischugrisch-samojedischen) Sprachfamilie*. Uppsala: Lundequistska.
1960 *Comparative Grammar of the Uralic Languages*. Stockholm: Almqvist & Wiksell.

Dolgopolsky, Aharon
1964 Gipoteza drevnejshego rodstva yazykov Severnoj Evrazii s veroyatnostnoj tochki zreniya [A probabilistic hypothesis concerning the oldest relationships among the languages of northern Eurasia]. *Voprosy yazykoznaniya* 2: 53–63.
1986 A probabilistic hypothesis concerning the oldest relationships among the language families of northern Eurasia. In: V. V. Shevoroshkin & T. L. Markey (eds.). *Typology, Relationship, and Time: A Collection of Papers on Language Change and Relationship by Soviet Linguists*, 27–50. Ann Arbor (MI): Karoma.

Dunn, Michael & Angela Terrill
2012 Assessing the lexical evidence for a Central Solomons Papuan family using the Oswalt Monte Carlo Test. *Diachronica* 29/1: 1–27.

Gamkrelidze, Thomas V. & Vjačeslav Vs. Ivanov
1995 *Indo-European and the Indo-Europeans*. 2 vols. Berlin / New York: Mouton de Gruyter [First publ. in Russian: *Indoevropejskij yazyk i indoevropejtsy*. 2 vols. Tbilisi: Tbilisi State University, 1984].

Helimski, Eugene
2000/2001 Early Indo-Uralic linguistic relationships: real kinship and imagined contacts. In: E. A. Helimski. *Komparativistika, uralistika: Lektsii i statyi* [*Historical Linguistics, Uralistics: Lectures and Papers*], 489–501. Moscow: Yazyki slavyanskikh kul'tur, 2000 [republ. in: *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations. Papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki 8–10 January, 1999* (Mémoires de la Société Finno-Ougrienne 242), 187–205. Helsinki, 2001].
2002 Zu den iranischen Lehnwörtern im Ungarischen. In: R. Blokland & C. Hasselblatt (eds.). *Finno-Ugrians and Indo-Europeans: Linguistic and Literary Contacts. Proceedings of the Symposium at the University of Groningen, November 22–24, 2001* (Studia Fenno-Ugrica Groningana 2), 105–111. Maastricht: Shaker.

Illich-Svitych, Vladislav Markovich
1967 Materialy k sravnitel'nomu slovaryu nostraticheskikh yazykov [Materials for a comparative dictionary of the Nostratic languages]. In: *Etimologiya 1965*, 321–373. Moscow: Nauka.
1971–1984 *Opyt sravneniya nostraticheskikh yazykov* [*An attempt at a comparison of the Nostratic languages*]. Ed. by V. A. Dybo. Vol. 1: *Vvedenie, sravnitel'nyj slovar'* (*b* — *ḳ*), Moscow: Nauka, 1971. Vol. 2: *Sravnitel'nyj slovar'* (*l* — *ʒ́*), Moscow: Nauka, 1976. Vol. 3: *Sravnitel'nyj slovar'* (*p* — *q*), Moscow: Nauka, 1984.

Ivanov, Vjačeslav Vs.
2001 Southern Anatolian and Northern Anatolian as separate Indo-European dialects and Anatolian as a late linguistic zone. In: R. Drews (ed.). *Greater Anatolian and the Indo-Hittite Family* (Journal of Indo-European Studies. Monograph Series 38), 131–183. Washington (D.C.): Institute for the Study of Man.
2009 K issledovaniyu otnoshenij mezhdu yazykami [Certain issues of current research on relations between languages]. *Journal of Language Relationship* 1: 1–12.

Justeson, J. S. & L. D. Stephens
1980 Chance cognation: a probabilistic model and decision procedure for historical inference. In: E. C. Traugott, R. Labrum & S. C. Shepherd (eds.). *Papers from the Fourth International Conference on Historical Linguistics, Stanford, March 26–30 1979*, 37–45. Amsterdam: Benjamins, 1980.

Kassian, Alexei
2013a On Forni's Basque–Indo-European hypothesis. *JIES* 41/1–2: 181–201.
2013b *The Lezgian linguistic group within the framework of the Global Lexicostatistical Database*. Talk given at the conference "Comparative-Historical Linguistics of the 21st Century: Issues and Perspectives", Moscow, March 20–22, 2013. Available: https://www.academia.edu/3040336/ [accessed 20.04.2014]
2014 Lexical matches between Sumerian and Hurro-Urartian: possible historical scenarios. *Cuneiform Digital Library Journal* 2014/4: 1–23. Available: http://cdli.ucla.edu/pubs/cdlj/2014/cdlj2014_004.html
2015 Towards a formal genealogical classification of the Lezgian languages (North Caucasus): testing various phylogenetic methods on lexical data. *PloS ONE*, 10(2): e0116950.

Kassian, Alexei, George Starostin, Anna Dybo, Vasiliy Chernov
2010 The Swadesh wordlist. An attempt at semantic specification. *Journal of Language Relationship* 4: 46–89.

Kassian, Alexei & Ilya Yakubovich
2013 Anatolijskie yazyki [Anatolian languages]. In: Y. B. Koryakov, A. A. Kibrik (eds.). *Yazyki mira: Reliktovye indoevropejskie yazyki*

Perednej i Tsentral'noj Azii [*Languages of the World: Relict Indo-European languages of Western and Central Asia*], 15–26. Moscow: Academia.

Kessler, Brett
2007 Word similarity metrics and multilateral comparison. In: *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*. Prague. P. 6–14.

Kessler, Brett & Annukka Lehtonen
2006 Multilateral comparison and significance testing of the Indo-Uralic question. In: P. Forster & C. Renfrew (eds.). *Phylogenetic methods and the prehistory of languages*, 33–42. Cambridge, UK: McDonald Institute for Archaeological Research.

Kloekhorst, Alwin
2008a Some Indo-Uralic aspects of Hittite. *JIES* 36/1–2: 88–95.
2008b *Etymological Dictionary of the Hittite Inherited Lexicon*. Leiden / Boston: Brill.

Koivulehto, Jorma
2001 The earliest contacts between Indo-European and Uralic speakers in the light of lexical loans. In: *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations. Papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki 8–10 January, 1999* (Mémoires de la Société Finno-Ougrienne 242), 235–263. Helsinki: Suomalias-Ugrilainen Seura.

Kortlandt, Frederik
2010 *Studies in Germanic, Indo-European and Indo-Uralic*. Amsterdam / New York: Rodopi.

Kümmel, Martin Joachim
2007 *Konsonantenwandel. Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vegleichende Rekonstruktion*. Wiesbaden: Reichert Verlag.

McMahon, April & Robert McMahon
2005 *Language Classification by Numbers*. Oxford: Oxford University Press.

Nakhleh, Luay, Tandy Warnow, Donald Ringe & Steven N. Evans
2005 A comparison of phylogenetic reconstruction methods on an IE dataset. *TPhS* 103: 171–192.

Napolskikh, Vladimir
1997 *Vvedenie v istoricheskuyu uralistiku* [*Introduction to historical Uralistics*]. Izhevsk: Udmurt Institute for History, Languages and Literature.

Oswalt, Robert
1970 The detection of remote linguistic relationships. *Computer Studies in the Humanities and Verbal Behavior* 3: 117–129.
1991 A method for assessing distant linguistic relationships. In: Sydney M. Lamb & E. Douglas Mitchell (eds.). *Sprung from some common source: Investigations into the prehistory of languages*, 389–404. Stanford: Stanford University Press.
1998 A probabilistic evaluation of North Eurasiatic Nostratic. In: J. C. Salmons & B. D. Joseph (eds.). *Nostratic: sifting the evidence*, 199–216. Amsterdam: Benjamins.

Patri, Sylvain
2009 La perception des consonnes hittites dans les langues étrangères au XIIIe siècle. *Zeitschrift für Assyriologie* 99: 87–126.

Rédei, Károly
1986 *Zu den indogermanisch-uralischen Sprachkontakten*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
1988–1991 *Uralisches Etymologisches Wörterbuch.* 3 vols. Budapest: Akadémiai Kiadó.

Ringe, Don
1992 *On Calculating the Factor of Chance in Language Comparison* (TAPS 82/1). Philadelphia: The American Philosophical Society.
1998 A probabilistic evaluation of Indo-Uralic. In: J. C. Salmons & B. D. Joseph (eds.). *Nostratic: Sifting the Evidence*, 153–197. Amsterdam: Benjamins.

Ringe, Don, Tandy Warnow & Ann Taylor
2002 Indo-European and computational cladistics. *TPhS* 100/1: 59–129.

Salminen, Tapani
2002 Problems in the taxonomy of the Uralic languages in the light of modern comparative studies. In: *Lingvisticheskij bespredel: sbornik statej k 70-letiyu A. I. Kuznetsovoj*, 44–55. Moscow: Izdatel'stvo Moskovskogo universiteta.

Sammallahti, Pekka
1988 Historical phonology of the Uralic languages with special reference to Samoyed, Ugric, and Permic. In: D. Sinor (ed.). *The Uralic languages. Description, history and foreign influences*, 478–554. Leiden / New York / Köln: Brill.

Schindler, Jochem
1975 Armenisch *erkn*, Griechisch ὀδύνη, Irisch *idu*. *Zeitschrift für vergleichende Sprachwissenschaft* 89: 53–65.

Sinor, Denis
1988 Introduction. In: D. Sinor (ed.). *The Uralic languages. Description, history and foreign influences*, xiii–xx. Leiden / New York / Köln: Brill.

Starostin, George
2008 *Making a Comparative Linguist out of your Computer: Problems and Achievements*. Presentation at the Santa Fe Institute, August 12, 2008. Available: http://starling.rinet.ru/Texts/computer.pdf [accessed 20.04.2014].
2010 Preliminary lexicostatistics as a basis for language classification: A new approach. *Journal of Language Relationship* 3: 79–116.
2013 *Yazyki Afriki. Opyt postroeniya leksikostatisticheskoj klassifikatsii* [*Languages of Africa: A new lexicostatistical classification*]. Vol. 1: *Metodologiya Kojsanskie yazyki* [*Methodology. Khoisan languages*]. Moscow: Yazyki slavyanskoj kul'tury.

Starostin, Sergei
1993/2007 Rabochaya sreda dlya lingvista [Linguist's workspace]. In: S. Starostin 2007: 481–496. [First publ. in: *Bazy dannykh po istorii Evrazii v srednie veka*, vyp. 2. Moscow: Institut vostokovedeniya RAN, 1993: 50–64. Republ.: *Gumanitarnye nauki i novye informatsionnye tekhnologii*. Moscow: RGGU, 1994: 7–23.]
2007 *Trudy po yazykoznaniyu* [*Works in Linguistics*]. Moscow: Yazyki slavyanskikh kul'tur.
2007a Opredelenie ustojchivosti bazisnoj leksiki [Defining the Stability of Basic Lexicon]. In: S. Starostin 2007: 827–839.

Stüber, Karin
1998 *The Historical Morphology of* n*-stems in Celtic*. Maynooth: National University of Ireland.

Syrjänen, K., T. Honkola, K. Korhonen, J. Lehtinen, O. Vesakoski, N. Wahlberg
2013 Shedding more light on language classification using basic vocabularies and phylogenetic methods. *Diachronica* 30/3: 323–352.

Thomason, Sarah Grey
2001 *Language Contact*. Edinburgh: Edinburgh University Press.

Thomason, Sarah Grey & Terrence Kaufman
1988 *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

Thomsen, Vilhelm
1869 *Den gotiske sprogklasses indflydelse på den finske. En sproghistorisk undersøgelse*. København: Gyldendal.

Turchin, Peter, Ilia Peiros & Murray Gell-Mann
2010 Analyzing genetic connections between languages by matching consonant classes. *Journal of Language Relationship* 3: 117–126.

Winter, Werner
1996 Lexical archaisms in the Tocharian languages. In: H. H. Hock (ed.). *Historical, Indo-European, and Lexicographical Studies: A festschrift for Ladislav Zgusta on the occasion of his 70th birthday*, 183–194. Berlin / New York: Mouton de Gruyter.

Zhivlov, Mikhail
2014 Studies in Uralic vocalism III. *Journal of Language Relationship* 12: 113–148.

Response to Kassian *et al.*, "Proto-Indo-European-Uralic comparison from the probabilistic point of view"

Don Ringe
University of Pennsylvania

It seems only fair to acknowledge that this paper reflects a significant advance in the long series of attempts to demonstrate a genetic relationship between Uralic and Indo-European (IE), which I shall call the Indo-Uralic (IU) hypothesis. Unfortunately the authors still have not made a convincing case for IU, because of several methodological shortcomings which I shall discuss in turn. I will also point out a potential objection that the authors have met successfully, since the authors themselves do not do so.

The authors are right to observe that previous work has not solved the problem (section 1), and their general principles (section 2.1) are unobjectionable. The same cannot be said of their use of "Dolgopolsky classes" (D-classes) to represent the consonants of the comparanda. Phonemes have the singular merit of being linguistically real: we know that native speakers use phonemic contrasts to distinguish lexemes and affixes from one another, and the overwhelming regularity of sound change *in phonemic terms* guarantees that regular sound correspondences are a reliable indicator of descent from a common ancestor. D-classes share neither of those properties. In fact, they actually privilege words that resemble one another by chance over historically real cognate sets, because the sounds of the former are phonetically similar by definition, whereas regular sound correspondences exist both between phonetically similar sounds and between sounds which are no longer similar. (That is why improbable-looking cognates that nevertheless exhibit regular sound correspondences, like Ancient Greek *ámathos* and English

sand, or German *Hals* and French *cou,* are so convincing: no one would have posited those relationships on the basis of superficial similarity.) The authors note that, of the seven comparative experiments reported in Kessler 2007, the two which (just barely) found an ostensible relationship between IE and Uralic were among those that used D-classes, but they do not quote Kessler's cautionary remarks. I quote them here:

> Dolgopolsky did not explain how he gathered the statistics upon which his classes are based. Since the classes were introduced in a paper designed to show that Indo-European and Uralic, among other families, are related to each other, it is possible that the statistics were informed at least in part by patterns he perceived between those language families. There is therefore some small cause to be concerned that Dolgopolsky classes may be … not completely unbiased with respect to the research question. (Kessler 2007: 12)

I would state those cautions much more emphatically than Kessler. But even if there is no bias of any kind, no one has demonstrated that D-classes are realistic enough units to be used as a basis for proof of non-obvious linguistic relationships. For that reason alone the authors' method does not, in fact, come closest to modelling traditional comparative linguistics (as suggested at the beginning of section 4).

The 50-item wordlist that the authors have used also requires some discussion. It was reasonable of them (or their predecessors) to try to construct a wordlist of maximally stable meanings; it is the specifics that raise questions. Perhaps the best way to approach those questions is to compare the authors' list with the 40-item list of Søren Wichmann,[1] also intended to be maximally stable, which is based on an astonishingly wide sample of languages. Since Wichmann's list is shorter, one would expect all the items on it to appear on the authors' list as

[1]Available at asjp.clld.org/parameters, link Meanings; the whole Swadesh 100-word list is given, and the 40 most stable words are in boldface. I am grateful to Gerhard Jaeger for the reference.

well. But no fewer than ten are missing: *person, fish, skin, knee, breast, see, come, path, mountain,* and *full* (in addition to *louse,* which the authors replaced with *liver* for reasons stated in their paper — both words are on Wichmann's list). The authors must have used a different sample of languages to construct their list, and it cannot possibly have been a larger sample than Wichmann's; but without knowing what languages were used we have no way of knowing whether the list could conceivably be biased in favor of IU.

The authors' observations on the phylogenies of the families involved are cogent, but in one respect they have not taken their own advice: they use putative protoforms which are actually attested in only one subgroup of a family (section 3, point (5), shortly before the table of comparative data). It has to be emphasized that such protoforms are not validated by the comparative method for Proto-Indo-European (PIE) or Proto-Uralic (PU). If one is trying to establish a controversial relationship on the basis of comparative data, the inclusion of such questionable forms weakens the argument considerably. Of course there are semantic slots for which no word is reconstructable for PIE or PU, and in those cases one might try using all the words reconstructable for first-order daughters because one can do no better; 'bird' appears to be a case in point for PU. But there are more than a few slots for which the authors adduce not only a word reconstructable for PU, but also one or more reconstructable only for various daughters, and the PIE list also includes a few such cases. If lexemes that are reconstructable only for subgroups of IE and Uralic are to be employed, it is important to use all that can be reconstructed up to a specified level in the cladistic tree; otherwise the danger that data have been selected because they seem likely to give a particular result becomes too great. Unfortunately the authors do not discuss these points.

On the other hand, the authors have succeeded in sidestepping a potential problem resulting from the use of multiple words in a single semantic slot. For instance, since the authors adduce three PIE forms and five PU forms for the slot 'black', they have given their algorithm *fifteen*

chances, instead of just one, to score an acceptable match for that slot, and 30 slots out of the 50 exhibit that property to one degree or another. If probabilities of cognation were being calculated directly, something would need to be done to account for those "extra chances"; the slots would have to be multiplied to take account of all possible cross-comparisons, for example, or only one comparandum might be selected at random for each run of the software. But because the authors have opted for a permutation method, and because exactly the same degree of freedom is allowed in every randomized comparison of the wordlists (as well as in the real binary comparison), their results will not be affected. I mention this because it might not be obvious to the general reader.

Some of the authors' data are problematic; in particular, the PIE forms that are used as input for their algorithm are not likely to satisfy most mainstream Indo-Europeanists. For instance, the authors are reluctant to reconstruct "laryngeal" consonants for PIE. But by this point in the development of the field a three-laryngeal reconstruction of PIE is accepted by virtually all German-speaking specialists, as well as throughout western Europe, Britain, and the USA. The odds that so large a community is mistaken about something so basic are virtually nil. On the other hand, the authors' acceptance of Ivanov's "*θ" in the reconstruction of two lexical items is almost certainly incorrect, if only because one of the roots, 'eye' / 'see', actually has descendants with initial *s- outside of Anatolian; the obvious example is the Germanic verb 'see' (Gothic *saíhvan,* etc.). I am not suggesting that competent specialists are not entitled to make their own judgments about reconstructions, but I am insisting that if one wants to convince skeptics one must employ data that the skeptics will accept, regardless of one's own opinions — at least if the skeptics are mainstream scientific linguists.

It seems possible that meeting all the above objections might not alter the authors' results much, but the way they have processed their data is a different matter. They assert (early in section 4) that only seven of the 50 word-pairs[2] are acceptable cognates for "Nostratic"

[2]Actually many more, see above.

because they are the only ones that exhibit identical CC-skeletons (reckoned in D-classes) in both protolanguages. Once again similarity is prioritized over nonsimilar but plausible cognation; other pairs that are equally plausible include 'nose' (Nuclear IE *nas- : Proto-Finno-Permic *näri), 'hand' (PIE *$g^{h\underline{y}}$es-$\underline{r}$: Proto-Finno-Ugric [PFU] *käti), and 'ear' (Nuclear IE *$h_2$ews- : Proto-Samoyed **kawi). It is true that requiring identity of consonants makes the results easier to assess (see below), but allowing strictly specified divergences can't be that much more difficult. But it turns out that the authors' assessment of "identity" is not straightforward. The PU lexeme meaning 'drink' might begin with a vowel or with *y-; the former alternative is chosen, apparently arbitrarily, so as to compare it with the vowel-initial (or possibly laryngeal-initial) PIE lexeme. Comparison of the words for 'hear', PIE *$k^{\underline{y}}$lew- and PFU[3] *kuwli-, shows that if they are related one item or the other must have undergone metathesis of two of its consonants; the authors take that for granted, without discussion, and assign a CC-skeleton [*KL*] to both items. In comparing PIE *me- 'me' and "*ti" 'thou' (actually *me and *te, enclitic accusatives) with PU *min and *tin, the authors segment off the *-n of the PU forms so as to produce monoconsonantal roots. In the Appendix (section 6) they say of *min, "Final **-n* is treated here as a suffix, since it is absent in the evidently related pronoun **me* 'we'." An identical statement, mutatis mutandis, is made for the second-person pronoun. But in that case the meanings of the monoconsonantal roots are '1st-person' and '2nd-person' respectively, and *-n is presumably a singular marker. Since the 1pl. pronoun is also part of the list, and since the contrast between singular and plural appears to depend on the presence or absence of *-n, *-n should be included in the analysis of the singular forms. In short, only three of the seven word-pairs are completely straightforward. It is difficult to avoid concluding that the data have been both cherry-picked and massaged. For the construction of an etymological dictionary that might be defensible, but for input to an automatic process that is

[3]Not PU, though the word is not doubly asterisked.

supposed to yield mathematically reliable results it is frankly indefensible, and it seems best to say so straightforwardly. These objections are serious enough to invalidate the authors' results.

If we nevertheless accept the authors' data as presented, we encounter another fatal shortcoming. The authors claim to have seven PIE–PU word-pairs which are best candidates for cognation, and they note that, depending on which version of D-classes one employs, seven or more such pairs appear in only 1.9% or in only 0.5% of randomized comparisons (section 4). But note that three of their seven best candidates are the monoconsonantal forms meaning 'me', 'thee', and 'who', which are assigned an *"H"* for their second consonant in the D-class system because they have no second consonant. The fact that they exhibit only one consonant each is not idiosyncratic to particular language families; on the contrary, pronoun roots very commonly contain a single consonant each. Thus those three items do not really meet the authors' two-consonant identity criterion. If we exclude them, only four remain, and as the authors show in Fig. 4 and Fig. 5, four or more such pairs appear in either 32% or 18% of randomized comparisons, again depending on which version of D-classes one uses. Such a result is comparable to results of earlier studies and actually suggests that the earlier studies were right: lexical resemblances between IE and Uralic are enough greater than average to create reasonable suspicions of genetic relationship, but not great enough for statistical proof. Finally, note that the only two pairs which are not subject to any potential objection are 'name' and 'water' — the two which nearly everyone agrees are the most plausible IU cognates if there really is a genetic relationship. A reasonable assessment of the authors' research is that, though their method might in principle be better than those of their predecessors (leaving aside the problems with the data, see above), they have effectively obtained comparable results.

But there is a larger question that needs to be emphasized because it is seldom discussed: what is a reasonable statistical threshold for proof of relationship in long-distance comparisons? Consider the following. There

are about 300 known families of languages (on Johanna Nichols' estimate), or perhaps about 350 (on Lyle Campbell's estimate), between which the comparative method, rigorously applied, has not yielded acceptable proof of relationship. Suppose that we take one language family as given and search for a pattern of similarities between it and all the others. If we find a pattern of similarity between our given family and one of the others which we calculate should appear by chance only once in 300 trials, *we are right in the middle of the chance range,* because 300 trials is (roughly) the number that we can make and we have succeeded only once. A probability of one in 300 is much smaller than what is usually accepted as probative in statistics, yet the logical structure of the problem shows that a threshold of one in 20 (5%) or even one in 100 (1%) is utterly inadequate for proof; we must demand probabilities *significantly smaller* than one in 300 to prove anything at all in this case. We do not know how much smaller they need to be, but it seems reasonable to accept Nichols' rule of thumb (see Nichols 1996): if we accept only probabilities about two orders of magnitude smaller than the middle of the chance range, we are unlikely to mistake a chance pattern for a historically real one. That means using a threshold of one in 30,000.

I can already hear the cries of outrage: who has ever used so stringent a criterion for linguistic relatedness? In fact the comparative method, if applied rigorously enough, routinely yields results that probative or even moreso; see Ringe and Eska 2013: 265–75 for a demonstration based on brute-force calculation of probabilities in a single case (English and Tocharian B).[4] A second method that has achieved such results is Johanna Nichols' "individual-identifying threshold" (Nichols 1996) as applied to structured sets of grammatical morphemes; for a demonstration (using the Algonkian prefix paradigm) see Ringe and Eska 2013: 276–8. Of course it is possible to

[4]I here take the opportunity to correct some minor errors in that passage. On p. 266 the number of English–German cognates should be 83, and the English and German words for 'yellow' should be underlined on p. 269; on p. 267 French 'ear' should be ɔr̲ɛj̲. On p. 272, line 5, the text should read "fewer than one out of every 10,000." The other figures are correct.

demand a similar level of proof using the permutation method. The table in Kessler 2007: 12 shows that most of his experiments yielded *p*-values less than .001 (one in 1,000) for all the valid clades of IE and Uralic, and for each of those families as a whole; it would be surprising if at least some of those results did not clear the one-in-30,000 threshold.

Unfortunately proof of that order seems out of reach for IU. The comparative method, strictly applied, has yielded no such result, nor has any of the permutation tests that have been run. To apply Nichols' method one needs a structured set of grammatical morphemes that is either fairly large or fairly quirky, or else a tightly structured set of lexemes (such as numerals), that are attested in both (or all) the languages under consideration. I doubt that IE and Uralic will offer researchers any such construct, though of course an attempt to prove IU by that method should be made. It would also be worthwhile to try to use two of the methods together: for example, to obtain a result by Nichols' method for the grammatical morphemes (only) and a brute-force or permutation result for lexical comparison (only), and then, since the two results will obviously be independent, to multiply them and see whether the one-in-30,000 threshold can be reached. I'm not optimistic, but it should be tried.

Finally, a more general point about methodology should be made. It would be grossly unscientific to regard probabilistic tests as an extraordinary tool for use only in special circumstances — useful for convincing colleagues who just aren't "intuitive" enough to see "the obvious", but in no sense fundamental. On the contrary: the structure of reality is well described by mathematics; probabilistic testing of wordlist comparisons is telling us the truth about IU — namely, that potential cognates are too sparse to provide solid proof of relationship. That's not ever going to change. Hoping that in the future some colleague will come up with a new and unforeseen test that will suddenly, magically find the resemblances that current methods have overlooked isn't just overoptimistic; it's irrational.

I admit that I am deeply disappointed by this situation. Like many other Indo-Europeanists, I continue to suspect

that there really is a genetic relationship between IE and Uralic (though not necessarily between any other potential subgroups of "Nostratic"), and I will be happy to applaud anyone who can provide robust, unassailable statistical proof of such a relationship. But sober consideration of the facts over many years has convinced me that no such proof will ever be forthcoming. The evidence just isn't there, and it seems best to accept the disappointment and go work on something else.

References

Kessler, Brett
2007 Word similarity metrics and multilateral comparison, in Nerbonne, John, Mark T. Ellison, and Grzegorz Kondrak (eds), *Proceedings of the Ninth Meeting of the ACL Special Interest Group on Computational Morphology and Phonology (SIGMORPHON)*, 6-14. Stroudsburg, PA: Association for Computational Linguistics) Downloaded from spell.psychology.wustl.edu/bkessler.html .

Nichols, Johanna
1996 The comparative method as heuristic, in Durie, Mark, and Malcolm Ross (eds), *The Comparative Method Reviewed*, 39-71. Oxford, OUP.

Ringe, Don, and Joseph F. Eska
2013 *Historical Linguistics: Toward a Twenty-First Century Reintegration*. Cambridge, CUP.

Response to Kassian et al., "Proto–Indo-European–Uralic comparison from the probabilistic point of view."

Brett Kessler
Washington University in St. Louis

Those of us who are of a certain age learned long ago to distrust papers that contained Swadesh word lists or numbers. Either the linguistics or the mathematics tended to be so distorted that one never knew what to make of the results. But progress has been made, thanks in large part to Ringe's initiative beginning in the 1990s, and historical linguistics has now assembled a set of valid and useable lexicostatistical methodologies. This work of Kassian and colleagues makes good use of those relatively new technologies, which I first quickly summarize.

1. Lexicostatistics and Bias

Saussure's arbitraire du signe holds that the connection between signifiers – here, the pronunciation of root morphemes — and the concepts they signify is almost entirely arbitrary. It follows that if a linguist takes two unrelated languages and for an infinite number of concepts finds the roots that signify the same concept in those two languages, any comparison function will give the same answer for semantically matched roots (the two roots in Language A and Language B that signify the same concept) as for randomly paired roots (a root in Language A and a root in Language B that are paired without regard to their meaning). In a more finite list of, say, 50 concepts, chance variation will enter the picture: some data sets will show more similarity for the set of semantically matched roots, others for the set of randomly paired roots. If two languages are related, though, cognates comprise an additional source of similarity that may accrue disproportionately to semantically matched roots. That

opens up the possibility that related languages can be identified by comparing the overall similarity of semantically matched roots with that of randomly paired roots, if only we can get a handle on the underlying chance variation. Chance can never be defeated, but we can get a useful picture of it by running Monte Carlo permutation tests. By shuffling the roots (randomly re-pairing them without duplication) a million times and measuring the similarity of the concept lists for each reshuffling, we can get an excellent estimate of how likely it is that chance would give paired concepts such a good (or even better) similarity measure as we find among the original data set with the semantically matched roots. This amounts to a statistical significance test as used in all experimental sciences. It lets us boil down the evidence to a probability value *p*, which reflects the likelihood that the observed similarity of semantically matched roots is no better than we would expect to find if only pure Saussurean arbitrariness was at play. Small *p*'s are good indications that the languages are related, and the 0.005 that the authors report here is very indicative indeed. In fact, when I tried my hand at replicating their experiment using their data and procedures, I got 0.003, which is even a wee bit better. It is laudable that the authors made their data available and explained their methodology so well that such close replication was possible.

For better or worse, these methodologies form more of a research paradigm than a fixed recipe. Many publications that use them are really explorations of alternative methodological parameters rather than concerted efforts to test whether certain languages are related (e.g. Ringe 1992; Kessler 2001, 2007). This flexibility is good because different linguistic situations might call for different measures. For example, for some languages it might not be practical to follow Kassian and colleagues in first reconstructing the protolanguages nor to insist that every root have at least two non-glide consonants. The flexibility is bad because choice can permit the introduction of bias. I use the term to refer to decisions that can push the results of the analysis toward or away from the research hypothesis. As I discuss some of the details of the paper, I will frequently refer to possible

biases, even if they are small or speculative, because small bias tends to be a huge factor in the outcome of investigations of long-range relationships between languages. This happens in part because there is very little probative evidence to begin with, which lends an unfortunate amount of weight to even minutely biased decisions.

Of course, in investigating bias, I do not mean to be accusing the researchers of bad faith. Bias is almost always unintentional, and it often is due to imbalances in the literature and available data sets that are out of the researchers' control. Nostratic hypotheses in particular have been so popular that one always wonders whether published theories and reconstructions of constituent language clades might have been affected by them. If two Uralic reconstructions seem to be of comparable plausibility, might a Nostraticist have a bias toward publishing the one that seems most compatible with Indo-European? Unless she were specifically thinking of statistical significance testing, it would be irresponsible for her not to publish the solution that supports the broader Nostratic picture.

2. Concept List

The project logically starts with a concept list, which serves two key roles. First, it offers guidance as to which concepts are most likely to be represented by simple, native words that have been retained for a very long time. It might contain concepts like 'black' and 'water' rather than 'coffee' and 'butterfly'. Second, it protects against bias. Statistical tests are useless if researchers collect the 50 words that best demonstrate their hypotheses and ignore everything else. Unfortunately, these two goals cannot be reconciled perfectly. The Swadesh lists are overwhelmingly the best choice for avoiding charges of bias because of their high profile: *Swadesh list* is virtually a synonym for *word list*. There is also no reason to suspect that the Americanist Morris Swadesh biased his list to support the Indo-Uralic hypothesis. Unfortunately, the Swadesh 200 list was composed on horseback and the subsequent Swadesh 100 list in an armchair: the result of much experience and thoughtful consideration, but no

quantitative research on retention rates. The list used in this paper, from G. Starostin (2010), compromises by starting with the Swadesh 100, then ranking the concepts by estimates of their retention rates and taking the top 50. This is not a bad idea, though it is slightly worrying from the standpoint of bias analysis that the first justification Starostin gives for the split is that the top half contains 17 of Sergei Starostin's proposed Indo-Uralic cognates whereas the bottom (rejected) half contained only nine (2010: 91). Also worrying is that there were exceptions to this 50% rule. Nine concepts from the top 50 were replaced with nine concepts from the lower 50 for reasons that Starostin declined to explain (2010: 93). What emerges is clearly a fine list of concepts, but again it is hard to rule out bias. For example, 'hear' was one of the words that were moved up the ranks to keep them in the word list. That turns out to be one of the seven key matches reported in this article: PIE [clew] with Finno-Ugric [kuwli].

3. Signifiers

Choosing the signifiers to express the concepts is usually the trickiest part of a lexicostatistical study. One of the nice aspects of this study is that protolanguage reconstructions were used for the PIE and PU signifiers, making it particularly interesting for many readers of *JIES*. This paper follows best practice in using only the root morphemes for the tests, even though Table 4 obfuscates that somewhat by also including suffixes (extensions) for many of the PIE entries; I assume these were stripped off before the similarities were actually measured.

The obvious advantage of using PIE and PU reconstructions is that those forms should be closer to any Indo-Uralic forms, therefore returning higher similarity measures among the cognates than would attested languages. The obvious disadvantage is that no one knows for sure what the actual PIE or PU protoforms were. Decisions such as rejecting the laryngeal theory could end up biasing the study if, for example, we believe that the PIE consonant [h_1] corresponds to zero in PU and would therefore always fail a straightforward consonantal

similarity metric.

Uncertainty in protolanguage reconstruction is here mitigated somewhat by allowing alternatives and variants among the signifiers, but that can also distract us from noting bias. Linguists who believe the PIE root for 'name' is [nom] would be outraged if a study selected [lom] as the root form. Those who believe in [lom] would be outraged to see [nom] selected. But because both forms are listed as alternatives, I predict much less outrage, in part because it is not obvious that the wrong form can do just as much damage when it is one of multiple choices as when it is the only choice. In this particular case, a large part of the success of this study hinges on the variant [nom]'s matching PU [nimi] – which seems reasonable to me, but perhaps not to a [lom] adherent. A case that does bother me is the PU reconstruction of 'drink' as alternatively [jɨɤi] or [ɨɣi]. The latter form is there to accommodate a Proto-Samoyedic reconstruction [ə]. Because I am not a Uralicist I will not say that the [j]-less PU reconstruction [ɨɣi] is almost certainly wrong. But we have a responsibility to be alert to the possibility that an author who is an expert on Anatolian languages might have subconsciously found [ɨɣi] appealing because it reminded him of Hittite [eku], which happens to be ultimately responsible for one of the seven matches in this study (as PIE [eːɣg^{wh}]).

The authors emphasize that their selection of roots to project back onto the PIE and PU protolanguages is constrained by topology. This sounds like a great idea, but I have found it impossible to evaluate. My interpretation is that they believe that all the non-Anatolian languages form a clade. Therefore a root found everywhere but in Anatolian could be an apomorphy within that non-Anatolian clade, making it dangerous to project back onto PIE. Conversely, a root found in Anatolian and just one other IE language is secure as a PIE root. The same applies, mutatis mutandis, to the status of Samoyedic versus Finno-Ugric within Uralic. However, the authors also project back to the protolanguage non-Anatolian and non-Samoyedic forms in a variety of situations that are not clearly spelled out. Without further detail, it is difficult to say whether this degree of discretion is overly conducive to bias.

Another problem that confronts the vocabulary selector is that certain types of signifiers are not subject to Saussurean arbitrariness. Onomatopoeia may differ between cultures, but onomatopoetic words are nevertheless more likely to be more similar to each other than truly arbitrary words are. Including them may increase the amount of similarity for semantically matched roots, which would bias toward concluding that the languages are related. Therefore we are obliged to remove words on even moderate suspicion of onomatopoeia. Starostin's trimmed Swadesh list suffers from potential onomatopoeia much less than the original, but some signifiers that might fall in that category are: words for 'bird' that sound like the call of crows ([$h_2$aw]), words for 'foot' that sound like the pitter-patter of little feet ([ped]), words for 'mouth' that display rounded lips ([h_3oː-s], [sʲuwi]), words for 'nose' with nasal sounds ([nas], [næri]), and words for 'tongue' that look like [lal]. The onomatopoeia with the most impact in this study is words for 'drink' that end in a velar sound, the velar standing in for the epiglottal sound of swallowing (cf. the list of drinking verbs ending in a velar or uvular sound in Greenberg & Ruhlen 1992, there presented as proof of Amerind, with a *p* value of 0.0000000001024). In this study both the PIE and the PU reconstructions are *glug* words and are close enough that they are counted as one of the seven matches.

A more subtle violation of arbitrariness is in the distinction between lexical words and grammatical words. Grammatical words tend to be shorter than lexical words and may have simpler phonology, with fewer marked sounds. Therefore in a word list, similarity between two semantically matched grammatical roots, like [me] and [mi] 'I', will on average be higher than similarity between a grammatical root and a lexical root, such as [me] versus [sʲæðʲæ] 'heart'. That increase in the relative similarity between semantically matched words over the randomly paired roots masquerades as our indicator of genetic relationship. Therefore lexical and grammatical words must not be included in the same concept list. It is simplest to remove the grammatical words because they are in the minority. Concepts in this study which fall in that category

include 'I', 'we', 'thou', 'what', 'who', 'not', and probably 'one', three of which are responsible for a sizable fraction of the matches reported in this study. Having to remove them is unfortunate because they otherwise have highly desirable retention rates crosslinguistically, but it can't be helped unless the experiment is significantly restructured.

4. Similarity Metric

The similarity metric is the next component that invites evaluation. I imagine many readers will object that counting recurrent sound correspondences would be preferable to any approach involving similarity. It is easy to count up benefits of using sound correspondences. It would be more traditional, which would be more compatible with the authors' claim that their techniques are more traditional than previous evaluations of Indo-Uralic, which include Ringe (1998), who used sound correspondences. It can sometimes find highly divergent sound reflexes, which is a good thing, contrary to the authors' dismissing the utility of finding matches like [d] ≅ [p]. It can also make it easier to identify loans, which are a major problem in many phylogenetic studies. But similarity measures are also perfectly valid, provided they are applied objectively and with statistical controls for chance, as in the paper under discussion.

The paper's similarity metric requires the first two consonants of the two roots to have matching Dolgopolsky classes. The Dolgopolsky classes essentially group together consonants that have similar places of articulation and sonority, as described in section 2.2. They constitute a perfectly coherent way of classifying similarity, though they do not seem particularly informed by what we know about sound change: little attempt is made to group together sounds with the weaker (usually more sonorous) sounds they typically evolve into. More importantly, they raise questions of bias. Might Dolgopolsky, an ardent Nostraticist, have grouped together consonants that he believed correspond to each other across the constituent families, including IE and Uralic? Even more suspicions of bias are raised by the fact that the definition of Dolgopolsky classes keeps changing. The version I am

familiar with (Baxter & Manaster Ramer 2000: 175), for example, groups sibilant affricates like [tʃ] with velars like [k] – a rare instance of grouping together sounds that are especially likely to be related by common sound changes. The version the paper references (from Starostin's *Global Lexicostatistical Database*) groups them with sibilant fricatives like [ʃ]. The authors used the latter version, then pulled the affricates out into a group of their own and ran a second analysis which gave better *p* values. I didn't find any explanation of all these changes in the paper. In the right context, this sort of tinkering with the similarity metric would be praised as part of a meritorious process of improving and exploring the evolving lexicostatistical methodology. But in a statistical assessment of a particular theory, it could raise the suspicion that bias entered the process, or even that multiple versions of a test were run until it gave results that best supported the research hypothesis – a procedure that makes the resultant *p* values invalid unless special adjustments are made.

My main quibble with the paper's similarity metric is that requiring a close match of two consonants is too strict. The authors point out that all the roots that their metric rates as similar have struck Nostraticists as being cognate, such as 'water': PIE [wed] and PU [weti]. But it is not the job of the comparison function to identify only pairs that would on their own convince readers that the languages are related. It just needs to compute differential levels of similarity and leave it to the Monte Carlo test to estimate the cumulative significance; small amounts of similarity can add up. My own experience has been that testing with sequences of phonemes is at best a disappointment, and at worst a tragic source of bias in favor of language relatedness (Kessler 2001: 158–174). A disappointment because the second consonant tends to be subject to much stronger forces for lenition and assimilation than the first consonant, so that it mostly detracts from the evidence afforded by the first consonant (cf. Ringe's experimental results, summarized in 1.1). Tragic because a multiphoneme criterion makes it very easy to effectively encode root length into the analysis. In this paper, if a root has only one consonant, the metric takes into

consideration adjacent non-consonantal segments, which are otherwise ignored. Thus a form like [me] 'I' gets encoded as MH and [e:g^{wh}] 'drink' as HK, where K is any velar obstruent and H is technically a vowel but, for practical purposes, functions as an indicator that the root is short. The upshot is that the explicit coding of short roots makes them an order of magnitude easier to match with short roots in the other language. This enables the manifestation of the problem discussed earlier, where inclusion of grammatical roots like [me] adds similarity that heavily biases the significance testing toward concluding there is a historical connexion. I tried rerunning the replication I mentioned earlier, but not counting two roots as matching unless they really matched on two real consonants. This test gave a p value of 0.06, or 0.03 if I allowed initial vowels to count as a consonant. Such values are still highly suggestive but an order of magnitude less certain than the published findings. Of course I don't adduce these calculations as actual research results but as a quick and dirty illustration that neglecting this exception to l'arbitraire du signe can have a very great effect.

5. Monte Carlo Permutation

The last part of the experiment to evaluate is the Monte Carlo test. As far as I could tell, this was perfectly executed. The million rearrangements that were used are more than sufficient for a very precise estimation of p. A small point to think about, though, is the statistical handling of multiple variants. The paper handled this by trying all combinations of one variant from PIE with one variant from PU and picking the best score. Thus 'ashes' had one chance of matching (only one form for each language) whereas 'bird' had nine chances (three forms in each language). Because that same process was used for all rearrangements of the data during the significance tests, the advantage that accrues to 'bird' pretty much evens out, and I didn't spot any actual problems. However, if the situation arose that words of some specific part of speech or semantic category systematically had more variants than another, that algorithm would exaggerate their similarity with respect to randomized pairs. I could imagine that

causing a problem.

6. Conclusion

One advantage of having a binary similarity metric – one that says that roots either are or are not similar – is that it is easy to see which concepts contribute most heavily to the results: the seven listed in section 4. Let me conclude by summarizing how secure they are in terms of my above comments about methodology and bias.

- 'Water' [wed]≅[weti] is an excellent match.
- 'Name' is very good, just downrated a little by the difficulty of figuring out what the best PIE reconstruction is. It was the reconstruction [nom] that was responsible for the match with [nimi], and a strong case can be made for that reconstruction.
- 'Hear' sounds great – [clew]≅[kuwli] – but is somewhat tarnished by a couple of facts. One is that Starostin put it in his list of 50 concepts even though it failed his ranking criterion and even though he said he thought it best to exclude verbs: a possible source of bias. Another smaller problem is the topological fact that reflexes of these roots do not appear in Anatolian or Samoyedic. This could make us worry that the PIE and PU protoforms are wrong, or even that there may have been a loan between branches of IE and Uralic that were in closer contact than Anatolian and Samoyedic.
- 'Drink' is scary. The forms [e:g^{wh}]≅[iɣi] match in part on the sigil H, signifying no initial consonant. This word would be excluded by a strict criterion requiring two consonants, which makes one wonder if there was some bias involved in introducing the special coding rule that effectively treats root-initial vowels like consonants. Second, the PU form is difficult to explain; the variant [jiɣi], which wouldn't match [e:g^{wh}], seems more likely. Another problem is that both words are likely to be onomatopoeia, which would violate the arbitrariness doctrine on which the statistics are based.

- The short grammatical forms for 'I', 'thou', and 'who' are completely out of place here. They do not actually have two matching consonants, but rather have a special symbol that encodes the fact that they are short words. Since in both languages short grammatical words contrast with longer lexical words, this procedure violates the arbitrariness doctrine on which the statistics are based.

This leaves us with two good matches – not bad, considering how stringent the two-character matching criterion is, but probably not good enough to attain the desired 0.05 *p* cutoff.

As a final word, I hope this close and sometimes critical analysis will not be taken as anything other than intense scholarly interest in the details of one of the best mathematical assessments of the Indo-Uralic hypothesis that I have seen.

References

Greenberg, Joseph H. and Merritt Ruhlen
1992 Linguistic origins of Native Americans. *Scientific American* 267(5): 94–99.

Kessler, Brett
2001 *The Significance of Word Lists: Statistical Tests for Investigating Historical Connections Between Languages*. Stanford, CA: CSLI Publications.

Nugae Indo-Uralicae

Petri Kallio
University of Helsinki

The Indo-Uralic hypothesis (i.e. the idea of a genetic relationship between the Indo-European and Uralic language families) is often credited to the Danish linguist Vilhelm Thomsen (1869), though he was only the first to formulate it in a critical scholarly way (cf. even earlier but most amateurish proposals by Europaeus 1853, 1863). Ever since him Indo-Uralic has been popular in Scandinavia among both Indo-Europeanists (e.g. Pedersen 1933) and Uralicists (e.g. Collinder 1934), and even the term *Indo-Uralisch* was first introduced by a Swede (Sköld 1926). Especially in Denmark this unbroken tradition is carried on even today (see e.g. Hyllested 2008).

Russia is not a newcomer in this field either, even though its earliest pioneers were from the non-Russian parts of the empire, such as Finland (e.g. Europaeus above) and Estonia (e.g. Anderson 1879). However, the tradition broke during the second quarter of the 20th century when the Soviet authorities favored N. Ya. Marr's blatantly anti-genetic "Japhetic theory". After Stalin himself had rehabilitated genetic linguistics in 1950, the new generation of the Soviet comparativists emerged by the 1960s when the Indo-Uralic hypothesis "arguably reached its highest degree of etymological elaboration in V. M. Illich-Svitych's work on Nostratic linguistics", as we can read in the article by his three compatriots, Alexei Kassian, Mikhail Zhivlov, and George Starostin (henceforth Kassian & al.). While I do agree with them that, even as unfinished, Illich-Svitych's Nostratic etymological dictionary (1971–1984) is more solid than its updates (e.g. Bomhard 2008; Dolgopolsky 2008), his work does not really stand up to closer scrutiny either (see especially Campbell 1998).

Suffice it to say that the Indo-Uralic hypothesis remains far from being proved, which does not mean that it would be improbable, let alone impossible. Rather than the standard comparative method the most popular tool has recently been the probabilistic approach that is now also applied by Kassian & al. To put it crudely, all this means that instead of comparing two reconstructed proto-languages one only compares two 50-item or 100-item wordlists. Call me a grumpy old man, but I fail to see how such laziness would be progress, although it is of course faster and easier, requiring hardly any knowledge of the languages themselves. It is therefore no wonder that limited wordlists are particularly popular among non-linguists applying phylogenetic methods to languages (see now Pereltsvaig & Lewis 2015 for an excellent review). Needless to say, my cynicism is not directed at Kassian & al. who are all professional linguists, no matter if we disagree on some details.

It was Voltaire who allegedly stated, "L'étymologie est une science où les voyelles ne font rien, et les consonnes fort peu de chose" ["Etymology is a science in which vowels signify nothing at all, and consonants very little"] (Müller 1864: 238). Similarly, Kassian & al. neglect vowels and focus on consonant classes rather than consonant phonemes, something that no one would ever get away with in micro-level/short-range comparison. True, the reconstruction of Indo-European ablaut and vowel-coloring laryngeals would explain away almost any irregular vowel correspondence, but then again Kassian & al. now have no vowel-coloring laryngeal at all but only the velar fricative **x* (> Hittite etc. *h*). Although they do "hold somewhat different opinions on the subject", I personally think that the most recent attempt to revive monolaryngealism (Pyysalo 2013) has already demonstrated what a bad idea it is.

Leaving aside certain novelties by Kassian & al. (e.g. PIE *θ > Hittite *s*, Luwian *t*, Nuclear IE Ø), their Indo-European reconstructions in general look somewhat more old-fashioned than their Uralic ones, which is almost never the case in this field. Instead, the main obstacle to progress has long been the fact that in the case of Uralic

most people rely on English-language sources that are either obsolete (e.g. Collinder 1955–1960) or abysmal (e.g. Décsy 1990, 2006). This time, however, it is not difficult to guess that Zhivlov, one of the best contemporary specialists of Uralic historical phonology, has made sure that there is nothing fatally wrong on the Uralic side. This being said, I could complain about their internal classification of Uralic, which they say "is widely accepted by Uralists, see, e.g., Collinder 1960: 11, Sinor 1988, Sammallahti 1988, Napolskikh 1997: 256 ff." Yet the first two have already passed away, the third remains open to better ideas (p.c.), and the fourth is an ethnohistorian rather than a historical linguist (see Kallio 2015: 79–86 for my most recent discussion on Uralic dispersal).

On the other hand, even if one agreed with the traditional classification that during the fourth millennium BC Uralic split up into Finno-Ugric and Samoyed, the third millennium BC Finno-Ugric words and the first millennium BC Samoyed words should not be considered equal candidates for the status of the Uralic proto-term. Note that giving less weight to Samoyed would not at all change the statistics because, as Kassian & al. point out themselves, their resulting Indo-Uralic candidates include no "Uralic" words exclusively attested in Samoyed (or any other single subgroup for that matter). Now speaking of the results, their 50-item wordlist includes as few as seven Indo-Uralic candidates, four of which are short forms (Campbell & Poser 2008: 200, 249, 375–376). Moreover, some of the rest would look far less convincing without their laryngeal-free Indo-European reconstructions. Consider the words for 'name': Uralic **nimi* is obviously closer to **nom-n̥* than **h₃neh₃-men-*, recently analyzed as a derivative (Kloekhorst 2008: 517–519). Yet even less convincing is the verb for 'to drink' (regardless of whether we prefer Indo-European **e·gʰʷ-* or **h₁egʷʰ-*), since its Indo-Uralicness is crucially based on their novel Uralic reconstruction **i̯γi-* with no initial glide, although the earlier idea that **j-* was simply lost in Samoyed is far more plausible than their new idea that **j-* appeared out of nothing elsewhere in Uralic.

Hereby Kassian & al. reach a startling conclusion that the lexical evidence for Indo-Uralic is even closer to zero

than I had previously thought. Interestingly, however, the remaining Indo-Uralic candidates include three pronouns, viz. 'I', 'thou', and 'who'. At least I would long ago have turned against Indo-Uralic if I had not been aware of the grammatical evidence which, most of all, involves paradigmaticity (see Čop 1975; Kortlandt 2010: 387–428). Note that what is offered as grammatical evidence should never lack paradigmaticity, since isolated affixes are of no value when establishing genetic relationships (Nicholls 1996). In the case of Indo-Uralic in particular, the Boppian tradition of comparative grammar seems to be the only way forward, because superficial comparisons of few basic words have already been made for centuries. Then again, even the 21th century comparative Uralic grammar is still nowhere near the 19th century comparative Indo-European grammar. Thus, there will be a lot of work to do on the Uralic side alone before seriously moving on to comparative Indo-Uralic grammar, something that already Thomsen (1869: 1–2) pointed out. Indeed, the circle is closed because three out of his four Indo-Uralic candidates, 'to hear', 'name', and 'water', still keep on being repeated by Kassian & al.

Incidentally, Thomsen's fourth and last Indo-Uralic candidate was the word for 'honey', which Kassian & al. now regard as an Indo-Iranian loanword in Finno-Ugric, as have done many others before them (cf. already Munkáczi 1901: 466–468). As Finno-Ugric **meti* (> Finnish *mesi*, Hungarian *méz*) looks more like Indo-European **med^{h}u* rather than Indo-Iranian **mad^{h}u* (> Sanskrit *mádhu*, Avestan *maδu*), the borrowing situation is supposed to have preceded the Indo-Iranian merger of Indo-European non-high vowels. Thus, one may only wonder on which grounds we even speak of Indo-Iranian because this merger, if any, was the defining phonological innovation of Indo-Iranian. Although there have also been other suggested sources such as Balto-Slavic and Tocharian (see e.g. Napolskikh 2001: 372), it seems much more likely to me that the Finno-Ugric word for 'honey' was borrowed simultaneously with the Finno-Ugric word for 'bee', **mekši* (> Finnish *mehi-*, Hungarian *méh*). Similarly, the source of the latter was not Indo-Iranian **makš-* (> Sanskrit *mákṣ(ik)ā*, Avestan

maxšī) but the preceding, yet post-*ruki*, proto-stage that can best be called dialectal Indo-European (Kortlandt 1994: 93–94; Beekes 1997: 25). The decision to call it Indo-Iranian and then to reject all the earlier contacts between Indo-European and Uralic seems tendentious (cf. Kallio 2002: 35–37).

Then again, Kassian & al. at least acknowledge the fact that there are dozens of loanwords which "reflect a profound Indo-Iranian influence on Finno-Ugric peoples, especially in the areas of agriculture and cattle-breeding". In their words, "the previous literature on this topic is enormous", so that it is indeed futile to question the existence of all such borrowings, especially if one does so without bothering to quote a single primary study (cf. Heggarty & Renfrew 2014: 1694–1695). Yet Kassian & al. are also more realistic than those who uncritically argue for hundreds of Indo-Iranian loanwords in Uralic (e.g. Katz 2003 whose *Frühurarisch* is once again closer to PIE than to PII, not to mention that also his *Ururalisch* with mobile accent and ablaut is almost closer to PIE than to PU reconstructed by any other Uralicist).

As far as the Indo-Uralic hypothesis is concerned, it is easily far more promising than most other hypotheses recently debated in this journal, since even its alleged opponents call it "plausible but inconclusive" (Campbell & Poser 2008: 162), telling us that "you can believe in it if you want" (Koivulehto 1993: 189). Indeed, there are only a few actual anti-Indo-Uralicists (e.g. Marcantonio 2014 who is also an anti-Indo-Europeanist and an anti-Uralicist), whereas most are simply skeptics who, however, are sometimes falsely called anti-Indo-Uralicists for straw man reasons. While I, too, still keep a wait-and-see attitude to Indo-Uralic, I could not agree more with Kassian & al. that "it is recommendable to search for a more appropriate explanation than chance coincidence".

References

Anderson, Nikolai
1879 *Studien zur Vergleichung der ugrofinnischen und indogermanischen Sprachen*. Dorpat: Laakman.

Beekes, Robert S. P.
1997 Historical phonology of Iranian. *The Journal of Indo-European Studies* 25: 1–26.

Bomhard, Allan R.
2008 *Reconstructing Proto-Nostratic: Comparative Phonology, Morphology, and Vocabulary*. Leiden: Brill.

Campbell, Lyle
1998 Nostratic: a personal assessment, in Joseph C. Salmons & Brian D. Joseph (eds.) *Nostratic: Sifting the Evidence*, 107–52. Amsterdam: Benjamins.

Campbell, Lyle & William J. Poser
2008 *Language Classification: History and Method*. Cambridge: Cambridge University Press.

Collinder, Björn
1934 *Indo-uralisches Sprachgut: die Urverwandtschaft zwischen der indoeuropäischen und der uralischen (finnischugrisch-samojedischen) Sprachfamilie*. Uppsala: Lundequistska bokhandeln.
1955–60 *A Handbook of the Uralic Languages*. Stockholm: Almqvist & Wiksell.

Čop, Bojan
1975 *Die indogermanische Deklination im Lichte der indouralischen vergleichenden Grammatik*. Ljubljana: Slovenska akademija znanosti in umetnosti.

Décsy, Gyula
1990 *The Uralic Protolanguage: A Comprehensive Reconstruction*. Bloomington (IN): Eurolingua.
2006 *The Common Finno-Ugric Language*. Bloomington (IN): Author House.

Dolgopolsky, Aharon
2008 *Nostratic Dictionary*. Cambridge: McDonald Institute for Archaeological Research.

Europaeus, D. E. D.
1853 *Komparativ framställning af de finsk-ungerska språkens räkneord, till bevis för Ungrarnes stamförvandtskap med Finnarne, och den indogermaniska folkstammens urförvandtskap med den finsk-ungerska*. Helsingfors: Frenckell.
1863 *Der Urstamm der indoeuropäischen Sprachenfamilie und seine vor-indoeuropäischen Abzwigungen, namentlich die finnisch-ungarische*. Helsingfors: Knauer.

Heggarty, Paul & Renfrew, Colin
2014 Western and Central Asia: Languages. Colin Renfrew & Paul Bahn (eds.) *The Cambridge World Prehistory*, 1678–1700.

Cambridge: Cambridge University Press.

Hyllested, Adam
2008 Internal reconstruction vs. external comparison: The case of the Indo-Uralic laryngeals, in Jens Elmegård Rasmussen & Thomas Olander (eds.) *Internal Reconstruction in Indo-European: Methods, Results, and Problems*, 111–136. Copenhagen: Museum Tusculanum.

Illich-Svitych, Vladislav Markovich
1971–84 *Opyt sravneniya nostraticheskikh yazykov* [*An Attempt at a Comparison of the Nostratic Languages*]. Moskva: Nauka.

Kallio, Petri
2002 Prehistoric contacts between Indo-European and Uralic, in Karlene Jones-Bley & al. (eds.) *Proceedings of the Thirteenth Annual UCLA Indo-European Conference*, 29–44. Washington (DC): Institute for the Study of Man.
2015 The language contact situation in prehistoric Northeastern Europe, in Robert Mailhammer & al. (eds.) *The Linguistic Roots of Europe*, 77–102. Copenhagen: Museum Tusculanum.

Katz, Hartmut
2003 *Studien zu den älteren indoiranischen Lehnwörtern in den uralischen Sprachen*. Heidelberg: Winter.

Kloekhorst, Alwin
2008 *Etymological Dictionary of the Hittite Inherited Lexicon*. Leiden: Brill.

Koivulehto, Jorma
1993 Suomen kielen indoeurooppalaiset lainasanat ja indoeurooppalais-uralilainen alkusukuongelma [Indo-European Loanwords in the Finnish Language and the Problem of the Genetic Relationship between Indo-European and Uralic]. *Academia Scientiarum Fennica: Year Book* 1992–1993: 181–190.

Kortlandt, Frederik
1994 From Proto-Indo-European to Slavic. *The Journal of Indo-European Studies* 22: 91–112.
2010 *Studies in Germanic, Indo-European and Indo-Uralic*. Amsterdam: Rodopi.

Marcantonio, Angela
2014 Uralic vs Indo-European contacts: Borrowing vs local emergence vs chance resemblances. *Journal of Estonian and Finno-Ugric Linguistics* 5/2: 29–50.

Müller, Max
1864 *Lectures on the Science of Language*. London: Longman, Green, Longman, Roberts, & Green.

Munkácsi, Bernát
1901 *Árja és kaukázusi elemek a finn-magyar nyelvekben* [*Aryan and Caucasian Elements in the Finno-Hungarian Languages*]. Budapest: Kiadja a Magyar Tudományos Akadémia.

Napolskikh, Vladimir
1997 *Vvedenie v istoricheskuyu uralistiku* [*Introduction to Historical Uralistics*]. Izhevsk: Udmurt Institute for History, Languages and Literature.
2001 Tocharisch-uralische Berührungen: Sprache und Archäologie, in Christian Carpelan & al. (eds.) *Early Contacts between Uralic and Indo-European*, 367–383. Helsinki: Société Finno-Ougrienne.

Nichols, Johanna
1996 The comparative method as Heuristic, in Mark Durie & Malcolm Ross (eds.) *The Comparative Method Revised*, 39–71. Oxford: Oxford University Press.

Pedersen, Holger
1933 Zur Frage nach der Urverwandtschaft des Indoeuropäischen mit dem Ugrofinnischen. In: *Liber semisaecularis Societatis Fenno-Ugricae*, 308–325. Helsinki: Société Finno-Ougrienne.

Pereltsvaig, Asya & Lewis, Martin W.
2015 *The Indo-European Controversy: Facts and Fallacies in Historical Linguistics*. Cambridge: Cambridge University Press.

Pyysalo, Jouna
2013 *System PIE: The Primary Phoneme Inventory and Sound Law System for Proto-Indo-European*. Helsinki: Unigrafia.

Sammallahti, Pekka
1988 Historical Phonology of the Uralic Languages with Special Reference to Samoyed, Ugric, and Permic, in Denis Sinor (ed.) *The Uralic Languages*, 478–554. Leiden: Brill.

Sinor, Denis
1988 Introduction, in Denis Sinor (ed.) *The Uralic Languages*, xiii–xx. Leiden: Brill.

Sköld, Hannes
1926 Indo-Uralisch. *Finnisch-Ugrische Forschungen* 18: 215–231.

Thomsen, Vilhelm
1869 *Den gotiske sprogklasses indflydelse på den finske: En sproghistorisk undersøgelse*. København: Gyldendal.

Lexicostatistics, Probability, and Other Matters

Alexei Kassian,[1] *Mikhail Zhivlov,*[2] and *George Starostin*[3]

First of all, we would like to express sincere gratitude to our colleagues Petri Kallio (PK), Brett Kessler (BK), and Don Ringe (DR), hence the *referees*, who were kind enough to read our paper on Indo-Uralic (Kassian, Zhivlov & Starostin 2015) with due care and suggest a number of comments, both favorable and critical. In the spirit of objectivity, we gratefully acknowledge a number of valuable additions and corrections suggested by the referees. However, it is also our opinion that the criticisms do not generally invalidate either our theoretical approach or the methods employed to realise it in practice.

Before proceeding, we would like to emphasize one general principle. Judging by the recent surge in publications in this field, it is rather obvious that a wide variety of mathematical algorithms and automated methods of analysis may be applied to linguistic data. However, it seems equally logical that it is only that particular method (or methods) that reflects the internal logic of the studied object and emulates the work of real experts in the field which will have an unquestionable advantage over the others. In other words, automated methods of analysis should not be seen as goals in themselves, but merely as tools that help us introduce higher standards of objectivity.

[1]Institute of Linguistics of the Russian Academy of Sciences / School for Advanced Studies in the Humanities of the Russian Presidential Academy of National Economy and Public Administration; a.kassian@gmail.com
[2]Department of Comparative Linguistics of the Russian State University for the Humanities / School for Advanced Studies in the Humanities of the Russian Presidential Academy of National Economy and Public Administration, zhivlov@gmail.com
[3]Department of the History and Philology of the Far East of the Russian State University for the Humanities / School for Advanced Studies in the Humanities of the Russian Presidential Academy of National Economy and Public Administration; gstarst1@gmail.com

With this consideration in mind, let us address the specific critical points that have been voiced in the discussion.

1. How our list of concepts was compiled. The list is primarily based upon the research of Sergei Starostin (2007a, see also G. Starostin 2013), who came up with a special procedure of calculating stability indices for 100 (actually, 110) Swadesh concepts. Starostin's sample comprises 110-item wordlists for 746 languages belonging to 14 language groupings, assembled within the framework of the *Tower of Babel* project (S. Starostin 1998–2005): 132 Sino-Tibetan lects, 99 Austro-Asiatic lects, 54 Altaic lects, 94 Austronesian lects, 36 Australian lects, 26 Khoisan lects, 33 North Caucasian lects, 21 Dravidian lects, 97 Indo-European lects, 7 Kartvelian lects, 69 Afrasian lects, 47 Tai lects, 17 Uralic lects, 14 Yeniseian lects. The "stability index" of each Swadesh item for each family was defined as M/L, where L = the number of languages in the family and M = the maximum number of languages within the taxon that use reflexations of the same root as the basic equivalent for the respective Swadesh meaning (e. g. the Slavic stability index for 'belly' is 0.38, since 5 out of 13 languages preserve reflexations of the same Proto-Slavic root **bryoːx-*). The results were then averaged to compile a "global" stability index.

Later, G. Starostin (2013) proposed to reduce the ranked 110-item wordlist to a 50-word subset, which is currently used in the GLD project for preliminary classification of poorly studied families (see G. Starostin 2013; 2014 for languages of Africa) and for purposes of long range comparison (where comparison of complete 110-item wordlists turns out to be cost-ineffective). This new 50-item wordlist comprises the most stable Swadesh concepts, according to S. Starostin's (2007a) stability indices, with the exception of 9 items (on which see the next section).

For comparative purposes, DR mentions the *Automated Similarity Judgment Program* 40-item wordlist (Wichmann et al. 2013), which, according to the authors of the ASJP project, represents the most stable words among the world's languages. DR mentions that this list "is based on an aston-

ishingly wide sample of languages" and should therefore be viewed as a highly reliable set of stable concepts. However, this seems to be a misunderstanding. It is true that the current version of the ASJP database covers a huge number of lects (almost 7000 40-item wordlists), but these 40 most stable concepts were themselves extracted from the complete 100-item list based on a relatively modest sample: 100-item wordlist data from only 245 languages were used in order to compile the current ASJP list of 40 concepts (Holman et al. 2008; for the actual list of 245 languages, see Brown et al. 2008).

Aside from the fact that the initial sample (compared to the 746 languages consulted by Sergei Starostin) is not so large, there is yet another problem with the ASJP wordlists: they differ quite seriously in quality. For instance, there seem to be 3 lexical errors and 19 erroneously transcribed forms within the 40-item ASJP wordlist for Modern Russian (see the linguistic supplement in Kushniarevich et al. 2015).

We suspect that these two factors are responsible for seriously skewed results of the entire ASJP project, and this is most transparently visible when it comes to average stability ranking: thus, according to Holman et al. 2008, the most stable concept on the list is 'louse' (!). From our point of view and according to the currently accumulated GLD evidence, this is an unexpected and unjustified ranking — it is true that the word 'louse' is relatively stable (#17 on S. Starostin's ranked list), but could it be more stable than, e.g., such personal pronouns as 'I' and 'we'?

2. Is our list of concepts biased? The currently accepted GLD 50-item list excludes nine out of S. Starostin's most stable concepts: 'this', 'that', 'liver', 'fish', 'neck', 'breast', 'full', 'to stand', 'to give', which were substituted by 'to kill', 'foot', 'horn', 'to hear', 'meat', 'egg', 'black', 'head', 'night', see G. Starostin 2013: 92–93 for reasons (that have nothing whatsoever to do with the Indo-Uralic hypothesis and were largely derived from a general analysis of the data from the 746-language sample) and discussion. For our IE-Uralic test, we took the liberty of reintroducing the high-ranking item 'liver' in the IE and Uralic wordlists instead of

'louse' only because a reliable phonological shape for the latter is non-reconstructible for Proto-IE. As a result of these substitutions, we have, indeed, unintentionally gained a new match: IE **k^{y}lew-* = Uralic **kuwli-* 'to hear'. But at the same time we have forfeited another match: IE **to-* = Uralic **te-* 'this' (the IE root is used for neuter gender and the oblique cases within the suppletive paradigm with **so-*).

Strictly speaking, it is possible that the replacement of 9 items could slightly bias our IE-Uralic comparison (as is fairly noted by BK), but we find it unlikely that this could significantly affect the obtained statistical results.

3. Do consonant classes relate to real life? DR states that historical linguists deal with regular sound correspondences and true etymological cognates, whereas consonant class analysis only reveals superficial similarities which are of little value, since they could arise simply due to chance; subsequently, in his opinion, we have no right to claim to be modeling the actual research of the historical linguist.

Since all of us are historical linguists and we have, among other lines of study, conducted some research on poorly studied linguistic families, we find this statement to be somewhat misleading. Any hypothesis of genetic relationship by its very nature begins with comparisons based on similarities; once a significant number of similarities has been registered by the researcher, he/she then begins to arrange them into regular patterns. It is theoretically possible, of course, to imagine a situation where *all* phonetic correspondences between two or more related languages would look like the textbook example of Old Indic *dva-* = Armenian *erku* '2', etc., but we know of no such situations in real life, and any proof of such a genetic relationship would cause a sensation in the linguistic world. Much more natural are situations where *some* regular phonetic correspondences are coupled with phonetic similarity and *some* are not: the comparative linguist first arrives at the former, aided with the "tool of similarity", then figures out the rest based on a more scrupulous analysis.

In general, historical phonology of a specific language should imply a bulk of typologically trivial phonetic corre-

spondences, and the majority of phonetic shifts should happen within the limits of phonetically justified consonant classes. Otherwise, if we only rely on regularity of sound correspondences ignoring phonetic typology, as suggested by DR, we can readily get a lot of curious results, e.g., "proof" that Basque is an IE language (cf. Kassian 2013).

Likewise, the goal of the automated method of consonant classes is to uncover *some* etymological cognates — not all of them, and not necessarily even the majority of them — and, moreover, it is highly probable that some of the uncovered pairs will indeed represent false cognates, i.e., chance phonetic coincidences. This, too, mimics the work of the comparative linguist, since it is well known that the first attempts at compiling etymological dictionaries for freshly studied families are often rife with etymological errors.

As an example, Ancient Greek *ámat^{h}o-s* 'sand' and English [sänd] are phonetically dissimilar, and therefore their etymological relationship cannot be revealed by primitive phonological analysis, as is justly noted by DR.[4] Nevertheless, even Modern Greek (Demotic) and Modern English (Received Pronunciation) demonstrate 4 *CC*-matches within the 50-item wordlist, and it should be noted that all four pairs are true etymological cognates:

1) Modern Greek *p ʹoð-i* = English [fut] 'foot' = *PT*
2) Modern Greek *n ʹe-o-* = English [nyuː] 'new' = *NH*
3) Modern Greek *ð ʹod-i* = English [tuːθ] 'tooth' = *TT*
4) Modern Greek *ð ʹio* = English [tuː] 'two' = *TH*[5]

[4]Strictly speaking, etymological relationship between *ámat^{h}o-s* and [sänd] is doubtful, and it is also unclear whether *ámat^{h}o-s* is indeed the basic term for 'sand' in at least some Ancient Greek dialects, but for the sake of convenience we follow DR and treat *ámat^{h}o-s* and [sänd] as true etymological matches within the Swadesh wordlist.

[5]Note that the permutation test between Modern Demotic Greek (Kassian 2011–2014) and modern literary UK English 50-item wordlists yields **significant** results, i.e., the probability *P* of getting at least four matches is < 0.05 (all conditions are the same as in Kassian et al. 2015): *P* = 0.020009 (2%) for basic GLD consonant classes; *P* = 0.012142 (1.2%) for more precise consonant classes.

4. Onomatopoeia, sound symbolism, phonosemantics. BK claims that some of our protoforms may be suspected to be onomatopoeic in origin, so that their phonetic recurrence in the IE and Uralic languages may be conditioned by their meanings rather than chance coincidence or etymological relationship. Among such items, BK quotes:

- IE **awi-* (**h₂ewi-*) ~ **xawi-* 'bird', sounds "like the call of crows" according to BK.
- IE **pod-* 'foot', sounds "like the pitter-patter of little feet".
- IE **oy-es* (**HoH-es-*), Uralic **sʸuwi* 'mouth', "display rounded lips".
- IE **nas-*, Uralic **näri* 'nose', "with nasal sounds".
- IE (Anatolian) **lal-* ~ **lol-* 'tongue' (the exact type of sound symbolism is not specified by BK).
- IE **eːg^{hw}-* (**h₁eg^{hw}-*), Ural **yɨγi-* ~ **iγi-* 'to drink', "the velar standing in for the epiglottal sound of swallowing"; this verb is especially important for us since it constitutes an IE-Uralic match.

The idea of sound symbolism of this kind seems somewhat odd to us. We believe that in general, onomatopoeic/symbolic explanations of the aforemen-tioned forms can be securely ruled out by two reasons.

1) If we resort to inventing *ad hoc* hypotheses about the non-arbitrariness of certain linguistic signs, we may reasonably eliminate quite a few lexical cognates in "short-range" comparison and thus bring into question the existence of conventionally accepted families. For instance, English [fut] will be disqualified as representing the sound of stamping, English [dɔg] will be disqualified for containing a velar sound which represents glottal barking, Russian *ruka* 'hand, arm' represents the sound of scratching, Russian *soncə* 'sun' contains a labial vowel that symbolizes the rotundity of the heavenly body, and so forth. Frankly, there is no set limit here for our fantasy, and there is no solid scientific basis that would allow us to separate the idea of "velar sounds standing in for the sound of swallowing" from the idea of "high vowels standing in for the idea of quickly performing an action", etc.

2) Wichmann et al. (2010) offer a rough statistics for 40 Swadesh items in about one half of the world's lan-

guages.[6] Since BK's 'to drink', 'nose', 'tongue' are included in the *ASJP* 40-item list of concepts, we can corroborate his suggestions of onomatopoeia with some actual evidence. Out of these three, only 'nose' appears to be susceptible to a non-chance phonetic shape. According to Wichmann et al. 2010: 854, the most normal shape of the words for 'nose' is *nani*. It is therefore tempting to explain the frequent occurrence of *n* in words for 'nose' across the world as sound symbolism, and perhaps this is indeed so. But what about the neighboring items which also seem to be affected by non-chance phonetic shapes: *naa* 'I', *nin* 'you (thou)', *nani* 'name', *nina* 'we' (Wichmann et al. 2010: 854)? We cannot propose any reasonable semantic, anatomic or behavioral explanation for the frequent occurrence of *n* in these words. Apparently one of the factors that should be taken into account is that *n* is simply the most frequently occurring consonant in the *ASJP* sample (see Wichmann et al. 2010: 848 for the table of sound frequencies). As to 'drink' and 'tongue', neither of these words shows any strong "onomatopoeic" signal according to Wichmann et al. 2010; ironically, statistically expectable shapes are *iaaa* (!) for 'to drink' and *aanaa* for 'tongue'.

Consequently, while we agree that the nasals in 'nose' could indeed be determined by sound symbolism (although cf. above), and possibly Hittite *lala-* 'tongue' could also represent a baby-talk term (*Lallwort*), we are unaware of any evidence in favor of a particularly high frequency of other sound patterns mentioned by PK among the world's languages.

5. Short function words. BK & DR notice that 3 out of our 7 IE-Uralic matches are pronouns which possess short *CV*-shapes, namely IE **me-* = Uralic **mi-n* 'I', IE **ti* = Uralic **ti-n* 'thou', IE **k^{w}i-* = Uralic **ku-* 'who'. According to BK & DR, such a predictable monoconsonantal shape can bias the results of the whole experiment.

Indeed, it is well known that personal and some other pronouns as well as negation exponents and other so-called function words tend to be short, i.e., monoconso-

[6]As noted above, statistical results based on *ASJP* data should be used with caution, but at least they can always be regarded as a starting point.

nantal or even monovocalic cross-linguistically. The question is how we should treat those of them that are included in the Swadesh list due to their historical relevance and high average stability.

The first option is to withdraw all function words from the comparative analysis altogether (as one of the referees does in his formal IE-Uralic comparison: Kessler & Lehtonen 2006, Kessler 2007). Under our approach, however, this is definitely a bad idea, since the goal of our study is to imitate the real work of historical linguists, and any historical linguist would say that at least such items as personal pronouns are not only *bona fide* objects of lexical comparison, but also ones that frequently help build up the most conclusive cases for language relationship.

The second option, since, indeed, the pronouns under our analysis automatically acquire the second consonant *H*, is to treat function words as true monoconsonantal items. This means that the shapes *CV* and *VC* will be mutually comparable, both of them being transcribed simply as *C*. E.g., Old Indic *ma:* 'I (me)' becomes equal to both Latin *me-* and Ancient Greek *em-* = [*M*], but, of course, not equal to Russian $m^{y}in^{y}a$ = [*MN*]. This is an acceptable solution, but we have intentionally rejected it in favor of an exceptionless approach, since normally we do not allow metatheses between *CV* and *VC* structures.

Nevertheless, if we do accept the monoconsonantal transcription for 'I', 'thou', 'we', 'what', 'who', and 'no' as described above, i.e., $C_1V = VC_1$ (e.g., IE $*k^{w}i-$ 'what' now matches both Uralic **ku-* 'who' and **ük:V* 'one', in case these happen to be compared during randomization), results of the permutation test between IE and Uralic 50-item wordlists remain **significant**, i.e., the probability *P* of getting at least seven matches is < 0.05 (all other conditions are the same as in Kassian et al. 2015):

- *P* = 0.036502 (3.7%) for basic GLD consonant classes;
- *P* = 0.011797 (1.2%) for more precise consonant classes.

The third option is to divide our 50-item wordlist into two subsets: content words and function words. The latter will then comprise only 6 items: the pronouns 'I', 'thou', 'we', 'what', 'who', and the negation exponent 'no'.

The subset of content words consists of 44 items with four IE-Uralic matches:

1) IE **kʲlew-* [*KL*] — Uralic **kuwli* [*KL*] 'to hear';
2) IE **nomn̥* [*NM*] — Uralic **nimi* [*NM*] 'name';
3) IE **wed-* [*WT*] — Uralic **weti* [*WT*] 'water';
4) IE **eːgʰʷ-* [*HK*] — Uralic **i̇γi-* [*HK*] 'to drink'.

The results of a permutation test between such truncated IE and Uralic 44-item wordlists become **insignificant**, i.e., the probability *P* of getting at least four matches is > 0.05 (all conditions are the same as in Kassian et al. 2015):

- *P* = 0.284015 (28.5%) for basic GLD consonant classes;
- *P* = 0.136144 (13.5%) for more precise consonant classes.

The subset of function words consists of 6 items ('I', 'thou', 'we', 'what', 'who', 'no') with three IE-Uralic matches:

1) IE **me-* [*MH*] — Uralic **mi-n* [*MH*] 'I';
2) IE **ti* [*TH*] — Uralic **ti-n* [*TH*] 'thou';
3) IE **kʷi-* [*KH*] — Uralic **ku-* [*KH*] 'who'.

The permutation test has returned **significant** results, i.e., the probability *P* of getting at least three matches is < 0.05 (all conditions are the same as in Kassian et al. 2015):

- *P* = 0.046292 (4.6%) for basic GLD consonant classes;
- *P* = 0.046013 (4.6%) for more precise consonant classes.

Finally, we can add two pronouns from the rest of the original Swadesh 100-item wordlist: 'this' and 'that'. The new 8-item function wordlist will then be as follows: Table 5.

Table 5. Proto-IE and Proto-Uralic 8-item function wordlists

#	Word	Indo-European	Uralic
42	**I**$_3$	*eg^{y}- (*h$_1$eg^{y}H) *me- ~ *əme- (*h$_1$me)	*mi-n
62	**Not**$_{30}$	*no ~ *ne	*e-
85	**That**$_{36}$	**ob^{h}o-	*to
86	**this**$_{28}$	*so- *to- *k^{y}o- ~ *k^{y}i-	*te
87	**thou**$_5$	*ti ~ *tu:-	*ti-n
95	**we**$_1$	*wey-s *n̥s-	*me
96	**what**$_{12}$	*k^{w}i-	*mi ~ *mɨ
98	**who**$_6$	*k^{w}i-	*ku-

Notes on the table: There are two main candidates for the PIE proximal deixis pronoun 'this' in the topologically "criss-crossed" configuration: (1) **k^{y}o- ~ *k^{y}i-*, meaning 'this' in Anatolian and Balto-Slavic, (2) the suppletive paradigm **so-* [nom. sg. masc. & fem.] / **to-* [nom.-acc. sg. neut., oblique, pl.], attested as 'this' in Tocharian, Germanic, Celtic and some other groups. We have to treat **k^{y}o- ~ *k^{y}i-*, **so-*, and **to-* as three synonyms.

The distal deixis pronoun 'that' is an unstable item in IE. The main candidate is ***ob^{h}o-*, attested, however, only in Anatolian. In other IE groups, the distal deixis pronoun 'that' looks like secondary formations, frequently based on the proximal pronoun **so- / *to-*.

In Uralic, demonstrative pronouns are more stable. The opposition **te* 'this' / **to* 'that' is attested in both Finno-Ugric and Samoyed branches and therefore can be reconstructed for Proto-Uralic.

The 8-item function wordlist includes four IE-Uralic matches:

1) IE **me-* [*MH*] — Uralic **mi-n* [*MH*] 'I';
2) IE **to-* [*TH*] — Uralic **te* [*TH*] 'this';
3) IE **ti* [*TH*] — Uralic **ti-n* [*TH*] 'thou';

4) IE *$k^{w}i$- [*KH*] — Uralic **ku*- [*KH*] 'who'.

The permutation test has again returned **significant** results, i.e., the probability *P* of getting at least four matches is < 0.05 (all conditions are the same as in Kassian et al. 2015):

- *P* = 0.019136 (1.9%) for basic GLD consonant classes;
- *P* = 0.019009 (1.9%) for more precise consonant classes.

The fourth option is to treat pronouns as normal Swadesh items, i.e., not to exclude them from the general wordlist (thus Kassian et al. 2015). We regard it as the best solution, since *volens nolens* pronouns are as much an integral part of historical linguistics as anything else.

Note that BK goes even further and suggests that any monoconsonantal roots (*CV* or *VC*) bias the result. We are not sure about this, since one of the goals of our algorithm is to emulate the work of historical linguists, whereas under traditional comparative approach monoconsonantal roots are certainly regarded as full-fledged *comparanda*. As an additional test, we accept the monoconsonantal transcription for all *CV* or *VC* roots in our IE and Uralic 50-item wordlists, i.e., $C_1V = VC_1$. In addition to the 7 IE-Uralic matches mentioned in Kassian et al. 2015, this gives us an 8th match: IE **ak*^{y}-*(moːn)* 'stone' = Uralic **kiwi* 'id.', because we treat non-initial glides as zero (NB: in reality we do not think that **ak*^{y}-*(moːn)* and **kiwi* are etymological cognates). Results of the permutation test between IE and Uralic 50-item wordlists remain **significant**, i.e., the probability *P* of getting at least eight matches is < 0.05 (all other conditions are the same as in Kassian et al. 2015):

- *P* = 0.025120 (2.5%) for basic GLD consonant classes;
- *P* = 0.007935 (0.8%) for more precise consonant classes.

6. Biased consonant classes? BK and DR stress that the consonant classes we use are intuitive and not based on typological statistics; in other words, either Dolgopolsky or we ourselves could intentionally compile consonant classes in such a way that IE-Uralic comparison would yield a certain number of phonetic matches.

It is true that consonant classes are not based on formal statistic data (as we have explicitly stated), since modern linguistics still lacks a representative universal database of diachronic sound shifts. Nevertheless, the idea that all sounds of the world's languages could be artificially distributed between classes just to make the IE-Uralic pairing look more "solid" is somewhat odd.

The basic idea behind consonant classes is as follows: sounds are arranged in classes in such a manner that for any given sound *X* its shift to sound *Y* is typologically more frequent than its shift to *Z* if *X* and *Y* belong to the same class and *Z* belongs to a different class. We assume that most historical linguists would agree that this condition works for the majority of pairs of consonant sounds within the GLD classes (Tab. 1 in Kassian et al. 2015). Naturally, there are exceptions and borderline cases, e.g., *v* is joined in one class with *p* (*P*-class) as opposed to the glide *w* (*W*-class) even though it is hard to understand whether the shift *v* ↔ *p* is more frequent than *v* ↔ *w* in the world's languages (in any case, *v* is not involved in our IE-Uralic comparisons).

The main question is how fine-grained one should strive to make the utilised classification of sounds, i.e. how many classes should be distinguished. Since we do not offer a definitive, universally applicable answer to this question, we offer two sets of classes: one called *basic GLD classes* (11 classes in total, Tab. 1 in Kassian et al. 2015) and an alternate, more fine-grained, set (13 classes in total, Tab. 2 in Kassian et al. 2015).

From an empirical standpoint, it is the place of articulation that turns out to be the most stable, i.e., the most expensive character for consonants. As one can see, the GLD classes are generally based on place of articulation. There is, however, at least one group of sounds that clearly violates this pattern — sibilant affricates (*c, č,* etc.), which often originate from velar stops (*k, g,* etc.), making the shift *k* > *č* typologically normal. The problem is that the shift in the opposite direction (*č* > *k*) is extremely rare; for the GLD classes, we assume that all sibilant affricates and fricatives should be joined together (*c, č* + *s, š*) as opposed to the velar/uvular *K*-class. An alternative, but also reason-

able approach was accepted by Dolgopolsky (1964, 1986), who includes sibilant affricates (*c, č*) into the *K*-class.

Dolgopolsky classes are listed in Tab. 6. Compared to the GLD, the main peculiarities of the Dolgopolsky classification are: sibilant affricates (*c ʒ č ǯ*…) are united with velars (*K*-class); initial *u-* is treated as a consonant of the *W*-class; initial *ŋ-* is treated as a consonant of the zero-class; lateral affricates (*ƛ*, i.e., the GLD *Q*-class) are not accounted for at all.

Table 6. Dolgopolsky consonant classes

P-class (labials): p b ɓ β f v…
T-class (dentals): t d ɗ θ ð…
S-class (sibilant fricatives): s z š ž…
Y-class (palatal glides): y…
W-class (labial glides): w ʍ…initial u-
M-class (labial nasals): m ɱ…
N-class (non-labial nasals): n ɳ ɲ ŋ…
R-class (liquida): r ɾ l ɭ ɬ ɮ…
K-class (velars, uvulars & sibilant fricatives): k g x ɣ q χ ʁ…, c ʒ č ǯ …
zero-class or H-class: ħ ʕ ʜ ʢ ʔ h ɦ ʔ and any vowels, initial ŋ.

Automated comparison between the IE and Uralic 50-item wordlists based on the Dolgopolsky classes has yielded the same 7 IE-Uralic matches as in the case of the GLD classes (all conditions are the same as in Kassian et al. 2015). The result of permutation test is **significant**, i.e., the probability *P* of getting at least seven matches is < 0.05:

- *P* = 0.014089 (1.4%) for Dolgopolsky consonant classes.

7. Multiple testing and the level of statistical significance. DR remarks that if we take the approximate number of language families all around the world to be *ca.* 300–350, then our IE-Uralic comparison should be regarded as a part of multiple testing and therefore the level of statistical significance should be lowered as compared with the commonly accepted value of 0.05.

It is true that whenever we conduct a series of tests on a dataset and want to be sure that our decision about

acceptance/rejection of the null hypothesis is correct *for each test*, we must proceed with the lower significance level. The most trivial (although not the optimal) technique is the Bonferroni correction, which implies that the accepted significance level is to be divided by the number of tests. So, when we consecutively compare the Proto-IE 50-item wordlist with the corresponding wordlists of the protolanguages of 300 families, applying the permutation test (300 pairwise tests in total), the accepted level 0.05 would mean that in 5% out of 300 cases our decision to reject the null hypothesis is expected to be wrong. Therefore, if we want to eliminate type-1 errors (i.e., erroneous rejections of the null hypothesis) in every test, we should accept the level 0.05 / 300, i.e., an extremely small value.

We believe, however, that it is really incorrect to assume that our IE-Uralic test is a part of multiple comparison of IE with all other families (and, correspondingly, Uralic with all other families). The hypothesis that Proto-Uralic is the closest relative of Proto-IE has a long tradition. It is based on various sorts of lexical and grammatical evidence. Although IE-Uralic etymological matches include those present in our 50-item wordlist, the complete set of traditionally proposed lexical IE-Uralic connections is much wider. Because of this, we believe that it is justified to restrict ourselves to just the IE-Uralic pair and not deal with the hypothetical results of comparison between IE and random protolanguages of another 300 families (it is indeed probable that due to chance, some of the 300 pairs would yield significant $P < 0.05$). Because of this, there is no need to introduce any correction in the method.

Another matter is that, from a certain point of view, the IE-Uralic comparison with different consonant classes can be regarded as multiple testing. As described above, we conducted three permutation tests of the IE-Uralic 50-item wordlists, using three sets of consonant classes: basic GLD classes (Tab. 1 in Kassian et al. 2015), more precise classes (Tab. 2 in Kassian et al. 2015), Dolgopolsky classes (Tab. 6 above). The obtained probabilities are: 0.018668, 0.004863, 0.014089. Accepting the 0.05 level and applying the Holm-Bonferroni correction (Holm 1979), which is suitable for such cases, we find that all the three obtained values are significant, i.e., the null hypothesis should be

rejected for every set of consonant classes.

Finally, in response to DR's skepticism concerning the capacity of probability tests to ever yield "solid proof of relationship": we may not have stressed this firmly enough in the original text, but the idea was not to devise the "ultimate" probability test that would finally provide such a proof. Instead, what we believe to have discovered (or at least confirmed) should rather be described as a "strong signal in favor of non-accidental similarities", which, upon further scrupulous analysis of said similarities, agrees better with a genetic than an areal interpretation. While there does exist a small chance of these similarities being accidental, or triggered by a very odd type of linguistic contact, the primary significance of the results of the test is that they justify both past and future research on "Indo-Uralic" (in such spheres as etymology, comparative morphology, explanation of anomalies and oddities in one family by means of comparison with the other, etc.) as a valid scientific hypothesis, fundamentally backed with statistic evidence, at least until such research happens to encounter some insurmountable obstacle (for instance, it is demonstrated that some other linguistic taxon shows an even stronger "relationship signal" with IE or Uralic than they do relative to each other, which, frankly speaking, is not highly likely in perspective).

8. "Laryngealistic" reconstruction for Proto-Indo-European. One of the referees' objections is that we do not adhere to the modern version of IE laryngealistic reconstruction or use it in an appropriate way. E.g., PK: "[…] Kassian & al. now have no vowel-coloring laryngeal at all but only the velar fricative **x* (> Hittite etc. *h*)". Similarly BK: "Decisions such as rejecting the laryngeal theory could end up biasing the study if […]". And finally DR: "[…] the authors are reluctant to reconstruct "laryngeal" consonants for PIE".

In response, we would like to point out that we use not one, but simultaneously two different approaches to Proto-Indo-European reconstruction, treating several forms as technical synonyms: the first one assumes the historical reality of the three laryngeals h_1, h_2, h_3 according to modern mainstream views, and the second one admits only the historical reality of the velar fricative *x*. Thus, e.g., for the

Proto-Indo-European entry 'ear' we technically reconstruct two synonyms: **ows-* (**h*$_2$*ews-* ~ **h*$_2$*ows-*) and ~ **xows-* (since the word is not attested in Anatolian, it could theoretically contain Early Proto-Indo-European **x-* (> cuneiform ‹ḫ›). It must also be stressed that there is no difference between the character strings **ows-*, **h*$_2$*ews-* and **h*$_2$*ows-* for our formal algorithm, which is why we write the laryngeal-containing forms in parentheses. Nor do we introduce a laryngeal-containing form if the morpheme in question is attested in Hittite with ‹ḫ, ḫḫ›, since there is strong evidence that the cuneiform ‹ḫ, ḫḫ› may only point to a Proto-Indo-European velar/uvular fricative.

References

Brown, Cecil H., Eric W. Holman, Søren Wichmann, Viveka Velupillai
2008 Automated classification of the world's languages: A description of the method and preliminary results. *STUF — Language Typology and Universals* 61.4: 285–308.

Dolgopolsky, Aharon
1964 Gipoteza drevnejshego rodstva yazykov Severnoj Evrazii s veroyatnostnoj tochki zreniya [A probabilistic hypothesis concerning the oldest relationships among the languages of northern Eurasia]. *Voprosy yazykoznaniya* 2: 53–63.
1986 A probabilistic hypothesis concerning the oldest relationships among the language families of northern Eurasia. In: V. V. Shevoroshkin & T. L. Markey (eds.). *Typology, Relationship, and Time: A Collection of Papers on Language Change and Relationship by Soviet Linguists*, 27–50. Ann Arbor (MI): Karoma.

Holm, Sture
1979 A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6.2: 65–70.

Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, Dik Bakker
2008 Explorations in automated language classification. *Folia Linguistica* 42.2: 331–354.

Kassian, Alexei
2011–2014 Annotated 100-item wordlists for the Greek group, Indo-European family. In: G. Starostin (ed.). *The Global Lexicostatistical Database*. Moscow/Santa Fe: Center for Comparative Studies at the Russian State University for the Humanities; Santa Fe Institute. Available: http://starling.rinet.ru/new100 [Accessed 07.05.2015].
2013 On Forni's Basque–Indo-European Hypothesis. *JIES* 41/1–2: 181–201.

Kassian, Alexei, Mikhail Zhivlov, George Starostin
2015 Proto-Indo-European-Uralic comparison from the probabilistic point of view. *JIES* 43(3-4) [the present issue].

Kessler, Brett
2007 Word similarity metrics and multilateral comparison. In: *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*. Prague. P. 6–14.

Kessler, Brett & Annukka Lehtonen
2006 Multilateral comparison and significance testing of the Indo-Uralic question. In: P. Forster & C. Renfrew (eds.). *Phylogenetic methods and the prehistory of languages*, 33–42. Cambridge, UK: McDonald Institute for Archaeological Research.

Kushniarevich, Alena, Olga Utevska, Marina Chuhryaeva, Anastasia Agdzhoyan, Khadizhat Dibirova, et al.
2015 Genetic heritage of the Balto-Slavic speaking populations: a synthesis of autosomal, mitochondrial and Y-chromosomal data. *PLoS ONE*, 10 (9): e0135820.

Starostin, George
2010 Preliminary lexicostatistics as a basis for language classification: A new approach. *Journal of Language Relationship* 3: 79–116.
2013 *Yazyki Afriki. Opyt postroeniya leksikostatisticheskoj klassifikatsii* [*Languages of Africa: A new lexicostatistical classification*]. Vol. 1: *Metodologiya Kojsanskie yazyki* [*Methodology. Khoisan languages*]. Moscow: Yazyki slavyanskoj kul'tury.
2014 *Yazyki Afriki. Opyt postroeniya leksikostatisticheskoj klassifikatsii* [*Languages of Africa: A new lexicostatistical classification*]. Vol. 2: *Vostochnosudanskie yazyki* [*East Sudanic languages*]. Moscow: Yazyki slavyanskoj kul'tury.

Starostin, Sergei
1998–2005 (ed.) *The Tower of Babel. An Etymological Database Project.* http://starling.rinet.ru/ [Accessed 07.05.2015]
2007 *Trudy po yazykoznaniyu* [*Works in Linguistics*]. Moscow: Yazyki slavyanskikh kul'tur.
2007a Opredelenie ustojchivosti bazisnoj leksiki [Defining the Stability of Basic Lexicon]. In: S. Starostin 2007: 827–839.

Wichmann, Søren, Eric W. Holman, Cecil H. Brown
2010 Sound Symbolism in Basic Vocabulary. *Entropy* 12(4): 844–858.

Wichmann, Søren, André Müller, Annkathrin Wett, Viveka Velupillai, Julia Bischoffberger, et al.
2013 The ASJP Database, version 16: http://asjp.clld.org/download [Accessed 07.05.2015].